

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 902.685 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-1  
Perfect score: 23  
Sequence: 1 tgaagaaggcagtagaagcttag 23

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.in.\*  
3: gb.env.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pr.\*  
9: gb.ro.\*  
10: gb.sts.\*  
11: gb.sy.\*  
12: gb.un.\*  
13: gb.vi.\*  
14: gb.htg.\*  
15: gb.pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	6	CS109170 Sequence
2	23	100.0	1731	6	C6647508 Sequence
3	23	100.0	1734	6	AX608629 Sequence
4	23	100.0	21781	1	AE014229 Streptoco
5	23	100.0	29072	6	AX602188 Sequence
6	23	100.0	167050	1	SAG766847
7	23	100.0	349980	6	C6655069 Sequence
8	23	100.0	349980	6	AX954529 Sequence
9	21	91.3	185291	14	AC123147 Rattus no
10	21	91.3	231447	14	AC123147 Rattus no
11	20.4	88.7	306616	14	AC095375 Rattus no
12	19.8	86.1	1378	6	AR353664 Sequence
13	19.8	86.1	1378	6	AX817901 Sequence
14	19.8	86.1	1379	6	AR353657 Sequence
15	19.8	86.1	1379	6	AR353663 Sequence
16	19.8	86.1	1379	6	AR353666 Sequence
17	19.8	86.1	1379	6	AX817894 Sequence
18	19.8	86.1	1379	6	AX817900 Sequence

19	19.8	86.1	1379	6	AX817903 Sequence
20	19.8	86.1	1384	6	AR353658 Sequence
21	19.8	86.1	1384	6	AR353660 Sequence
22	19.8	86.1	1384	6	AX817895 Sequence
23	19.8	86.1	1384	6	AX817897 Sequence
24	19.8	86.1	1385	6	AR353659 Sequence
25	19.8	86.1	1385	6	AX817896 Sequence
26	19.8	86.1	1390	6	AR353661 Sequence
27	19.8	86.1	1390	6	AX817898 Sequence
28	19.8	86.1	1393	6	AR353665 Sequence
29	19.8	86.1	1393	6	AX817902 Sequence
30	19.8	86.1	1447	6	AR353667 Sequence
31	19.8	86.1	1447	6	AX817904 Sequence
32	19.8	86.1	1731	6	C6647510 Sequence
33	19.8	86.1	1803	6	AR353656 Sequence
34	19.8	86.1	1803	6	AX817892 Sequence
35	19.8	86.1	10430	1	AE006575 Streptoco
36	19.8	86.1	11992	1	AE010058 Streptoco
37	19.8	86.1	51454	1	AE014157 Streptoco
38	19.8	86.1	110000	1	BA000034_07 Continuation (8 of
39	19.8	86.1	110000	1	BA000034_08 Continuation (9 of
40	19.8	86.1	110000	1	CP000003_10 Continuation (11 o
41	19.8	86.1	110000	1	CP000017_10 Continuation (11 o
42	19.8	86.1	110000	1	CP000056_11 Continuation (12 o
43	19.8	86.1	140072	14	AC133332 Felis cat
44	19.8	86.1	164494	9	AC154824 Mus muscu
45	19.8	86.1	270432	14	CT010447 Mus muscu
46	19.8	86.1	270432	14	CT010447 Mus muscu
47	19.8	86.1	319008	14	AC110439 Rattus no
48	18.8	81.7	60036	14	AC009610 Homo sapi
49	18.8	81.7	132003	8	AC138779 Homo sapi
50	18.8	81.7	164084	14	AC026580 Homo sapi
51	18.8	81.7	181895	14	CT025542 Mus muscu
52	18.8	81.7	191119	14	CT025542 Rattus no
53	18.8	81.7	200562	9	AC117689 Mus muscu
54	18.8	81.7	239438	14	AC107536 Rattus no
55	18.8	81.7	239808	14	AC128575 Rattus no
56	18.8	81.7	243280	14	AC128356 Rattus no
57	18.4	80.0	3310	8	HSL81824 Homo sapien
58	18.4	80.0	60634	14	AC009010 Homo sapi
59	18.4	80.0	60791	8	AC004749 Homo sapi
60	18.4	80.0	79041	8	AC009016 Homo sapi

## ALIGNMENTS

RESULT 1	CS109170	CS109170	Sequence 1 from Patent EP1541697.	23 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	CS109170	CS109170	GI:68148019				
DEFINITION	CS109170	CS109170	synthetic construct				
ACCESSION	CS109170	CS109170	other sequences; artificial sequences.				
VERSION	CS109170.1	GI:68148019	1				
KEYWORDS			Unl.J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.				
SOURCE			Detection of group B streptococcus				
ORGANISM			Patent: Bp 1541697-A 1 15-JUN-2005;				
REFERENCE			Mayo Foundation for Medical Education and Research (US)				
AUTHORS			Location/Qualifiers				
TITLE			1. .23				
JOURNAL			/organism="synthetic construct"				
FEATURES			/mol_type="unassigned DNA"				
source			/db_xref="taxon:32630"				
			/note="Oligonucleotide"				

## ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 1 TGAGAGGCAGTAGAAGCTTAG 23

RESULT 2
CO647508
LOCUS 1731 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 4465 from Patent WO0234771.
ACCESSION CO647508
VERSION CO647508.1 GI:41683480
KEYWORDS Streptococcus agalactiae
SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
ORGANISM Streptococcus.

REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 23; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 177 TGAGAGGCAGTAGAAGCTTAG 199

RESULT 3
AX608629
LOCUS 1734 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 6558 from Patent WO02092818.
ACCESSION AX608629
VERSION AX608629.1 GI:28404206
KEYWORDS Streptococcus agalactiae
SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
ORGANISM Streptococcus.

REFERENCE 1
AUTHORS Glaser,P., Rusniok,C., Chevallier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 6558 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1..1734
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 23; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 177 TGAGAGGCAGTAGAAGCTTAG 199

RESULT 4
AE014229
LOCUS 21781 bp DNA linear BCT 23-SEP-2002
DEFINITION Streptococcus agalactiae 2603V/R section 39 of 100 of the complete
genome.
ACCESSION AE014229
VERSION AE014229.1 GI:22533831
KEYWORDS Streptococcus agalactiae 2603V/R
SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
ORGANISM Streptococcus.

REFERENCE 1 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
12200547

REFERENCE 2 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
Direct Submission
Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1..21781
/organism="Streptococcus agalactiae 2603V/R"
/mol_type="genomic DNA"
/strain="2603V/R"
/serotype="V"
/db_xref="taxon:208435"
103..348
/gene="SAG0813"
103..348
/gene="SAG0813"
/feature="identified by Glimmer2; putative"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAW99700.1"
/db_xref="GI:22533832"
/translation="MNNQINHCFSIMVLGVTFQMKKEFTITGRMVARVNETLVCLCMV
SNLVDLMDIIEINIVFCVLVASTAGLEFCLEKERE"
387..674
/gene="SAG0814"
387..674
/gene="SAG0814"
/feature="identified by Glimmer2; putative"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAW99701.1"
/db_xref="GI:22533833"
/translation="MTYIGVPKEQKALSRGKLEKINGGKTLMSQMTQDLVGSTCM
IYLEDGSQTHPYEILKVDWDVQIRNRLKRSKEEVRIVRIHDIKELRPVVK"
719..934
/gene

```



```
/gene="SAG0815"
719..934
/gene="SAG0815"
/translation="identified by match to PFAM protein family HMM
PF01381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/CI family"
/protein_id="AAM99702.1"
/db_xref="GI:22533834"
/translation="MGLIKNRKLKELRARDQINGTELAKUTGVSRQTSILIERGEYTP
SVIAMKISQVFKELENVFLSVSEE"
934..1695
/gene="SAG0816"
934..1695
/gene="SAG0816"
/translation="similar to OMNI:NTL01SPL1089; identified by
sequence similarity; putative"
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99703.1"
/db_xref="GI:22533835"
/translation="MMKNFIIKQSDTTFVRIIKSLIGGFIGAILSGVGFIIFGQDK
YLSNIVQYFLWGRIVVITAFSLIYVIOIKYQKVFNVDSQSEIYRQINLR
HSGWTFYSISIVLNTLVNKLNIPODSTVLPIDLSLLFVLGLHYFLKVV
RNIRGKMTVATLRELKNNVLQDLAELESYKWCDFIVMNLGSFIFPTIYFVLFFI
SPVQKVEIATITTSIHIYILIKSLKAARHFYR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/translation="similar to OMNI:NTL01SPL0446; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99704.1"
/db_xref="GI:22533836"
/translation="MSKNNNTCLTETAFALAMALSMIPDFASWFTPSGAIPILL
FALRGTKYGLFAGLWGLHFLVSLGSQVFIYELAFISMGLAGVPSAKPKD
ALSSSKTKALSLAGLALTLRVYVHYIAGVIFWASYPKGMATLYLSLVNGTA
GLTLFFVVISIILVISYSPFLPKK"
complement(2584..3543)
/gene="nrdf-2"
/translation="SAG0818"
complement(2584..3543)
/gene="nrdf-2"
/translation="identified by match to PFAM protein family HMM
PF00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AAM99705.1"
/db_xref="GI:22533837"
/translation="MTTYEALNWNIEIDVDSKTEKLEQFWLDTRIPLNSLDDW
KLSAQEKDLGVGLTLDLTMOSETGEAIRADVRTPHEEVLNNIQLFQWESVHAK
RYSISFSTLNTKEIEBEFTWNNNEFLQEKARIINDIYANGNALQKRVASTYETFL
FVSGFTPLIYLNGLNANVAELIKLIRDESVMHTYIGYKFLAGNELPEDEQENFR
DMWYDILYOLYENEEKYTKTLVDGVGWTVEVMTFLRYNANKALMNLGQDPLPDTAND
VNPVNMGSTGTSHNDFESQVQNGYLLGSVEAMHDDDDNYGL"
complement(3746..5905)
/gene="nrdf-2"
/translation="SAG0819"
complement(3746..5905)
/gene="nrdf-2"
/translation="identified by match to PFAM protein family HMM
PF00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
/gene="SAG0815"
719..934
/gene="SAG0815"
/translation="identified by match to PFAM protein family HMM
PF01381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/CI family"
/protein_id="AAM99702.1"
/db_xref="GI:22533834"
/translation="MGLIKNRKLKELRARDQINGTELAKUTGVSRQTSILIERGEYTP
SVIAMKISQVFKELENVFLSVSEE"
934..1695
/gene="SAG0816"
934..1695
/gene="SAG0816"
/translation="similar to OMNI:NTL01SPL1089; identified by
sequence similarity; putative"
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99703.1"
/db_xref="GI:22533835"
/translation="MMKNFIIKQSDTTFVRIIKSLIGGFIGAILSGVGFIIFGQDK
YLSNIVQYFLWGRIVVITAFSLIYVIOIKYQKVFNVDSQSEIYRQINLR
HSGWTFYSISIVLNTLVNKLNIPODSTVLPIDLSLLFVLGLHYFLKVV
RNIRGKMTVATLRELKNNVLQDLAELESYKWCDFIVMNLGSFIFPTIYFVLFFI
SPVQKVEIATITTSIHIYILIKSLKAARHFYR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/translation="similar to OMNI:NTL01SPL0446; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99704.1"
/db_xref="GI:22533836"
/translation="MSKNNNTCLTETAFALAMALSMIPDFASWFTPSGAIPILL
FALRGTKYGLFAGLWGLHFLVSLGSQVFIYELAFISMGLAGVPSAKPKD
ALSSSKTKALSLAGLALTLRVYVHYIAGVIFWASYPKGMATLYLSLVNGTA
GLTLFFVVISIILVISYSPFLPKK"
complement(2584..3543)
/gene="nrdf-2"
/translation="SAG0818"
complement(2584..3543)
/gene="nrdf-2"
/translation="identified by match to PFAM protein family HMM
PF00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AAM99705.1"
/db_xref="GI:22533837"
/translation="MTTYEALNWNIEIDVDSKTEKLEQFWLDTRIPLNSLDDW
KLSAQEKDLGVGLTLDLTMOSETGEAIRADVRTPHEEVLNNIQLFQWESVHAK
RYSISFSTLNTKEIEBEFTWNNNEFLQEKARIINDIYANGNALQKRVASTYETFL
FVSGFTPLIYLNGLNANVAELIKLIRDESVMHTYIGYKFLAGNELPEDEQENFR
DMWYDILYOLYENEEKYTKTLVDGVGWTVEVMTFLRYNANKALMNLGQDPLPDTAND
VNPVNMGSTGTSHNDFESQVQNGYLLGSVEAMHDDDDNYGL"
complement(3746..5905)
/gene="nrdf-2"
/translation="SAG0819"
complement(3746..5905)
/gene="nrdf-2"
/translation="identified by match to PFAM protein family HMM
PF00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
/protein_id="AAM99706.1"
/db_xref="GI:22533838"
/translation="MSLKNIGDVSYFRLNNEINRPNVNGQIPLHDKKEALTAFKKENVQ
PNSKAPDSITDKIAYLLKYDYLEEAFNLKYPPEIEBLSTKLPDKKFRFKSFMAYKFK
YQYALKTNDGEVYLESIEDRVLFNALYFADGDEELATDLALEMISORYQAPATPSFLN
AGRRRGELVSCFLIQVTDNNALGRINSNALQLSRIGGGVIGISLNLREAGAPIKGP
AGAAAGVVPVNMKUFEDSFYSNQLGQAGVYDVFHPDIIISFLSTKKENADEKVR
VKTLSLGTIVPDKFELARNQEMYLFPYSIEREYGVFPYIDITIKYDELVANPNI
TKTINARDELETSIKQOESGYPYIINIDTANRTPVDGKIIMNSLCSLQVQKPS
LINDAQVLEMGTDISCNLGSVNLNMTSPDFGKSIKTMTALTFVTDSSNIENAVPT
IKNGAQAHPFELGAMGLSHVLAQNHYGSPESIEFTDIYFMLMNYWTLVLESNNIAR
ERQTTIVGFESKYADGTYPKYVSGKVFPSQDVKSLFANHFIEPKADWENLRVAVM
KQGLYHQNRLLAVAPNSIYINDCSAHPITQIRIERQEKKIGIKIYYPANGLATDTI
PYTTSAYDMRNVKSIYIIRTAATEHDQGLSMTLFLRSELPKELVWETKESQTTDRDLS
ILRNYAFNKGKVIYIIRTAATEHDQGLSMTLFLRSELPKELVWETKESQTTDRDLS
complement(5983..6207)
/gene="nrdfh"
/translation="SAG0820"
complement(5983..6207)
/gene="nrdfh"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2,
nrdfh-redoxin"
/protein_id="AAM99707.1"
/db_xref="GI:22533839"
/translation="MAAITVFSKNNCMQCKMTKPKLQDQADPBEINIDEKPEKIEYV
KNLGSAPVIEAGNVVFSQPSPKSLKELV"
6590..6853
/gene="ptsh"
/translation="SAG0821"
6590..6853
/gene="ptsh"
/translation="identified by match to PFAM protein family HMM
PF00381"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAM99708.1"
/db_xref="GI:22533840"
/translation="WASKDFHIVAEITGHARPATLLVOTASKFASDITLDYKGAIVNL
KSIKGMVSLGVGGQADVTISAGDADDAIAAIEBTMTKEGLA"
6858..8591
/gene="ptsi"
/translation="SAG0822"
6858..8591
/gene="ptsi"
/translation="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAM99709.1"
/db_xref="GI:22533841"
/translation="MTEMLKGIASDGVAVAKAYLLVQDLSFETVTVEDTNAEERL
DVALQSDLSVIREKAVEGEAAVFDALHMLVSDPENINGIKETIRAKQVNAE
TGLKVTOMFTIPFGMEDNPYMGRAADIRDVAKRVLAHLGLKLPNPATINEESIV
IAHDLTPSDTIAQLNKQFVKAFVYTNIGRSHSAIINARTLEIAAIVLGTNDITERVQDQG
```

Query Match 100.0%; Score 23; DB 1; Length 21781;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGAAGCGCAGTAGAAGCTTAG 23  
DB 7034 TCAGAAGCGCAGTAGAAGCTTAG 7056

RESULT 5  
AX602188  
LOCUS 29072 bp DNA linear PAT 17-FEB-2003  
DEFINITION Sequence 117 from Patent WO02092818.  
ACCESSION AX602188

VERSION	AX602188.1	GI:28402057	
KEYWORDS	Streptococcus agalactiae		
SOURCE	Streptococcus agalactiae		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1		
AUTHORS	Glaser, P., Rueniok, C., Chevalier, F., Frangeul, L., Lalioui, L., and Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieu-Cuot, P., and Kunst, F.		
TITLE	Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets		
JOURNAL	Patent: WO 02092818-A 117 21-NOV-2002; INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)		
FEATURES	Location/Qualifiers		
source	1..29072		
	/organism="Streptococcus agalactiae"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:1311"		
ORIGIN			
Query Match	100.0%;	Score 23;	DB 6; Length 29072;
Best Local Similarity	100.0%;	Pred. No. 2.6;	
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1 TGAGAGGCAGTAGAAGCTTAG 23		
Db	12072 TGAGAGGCAGTAGAAGCTTAG 12094		
RESULT 6			
SAG766847			
LOCUS	SAG766847	167050 bp	DNA linear BCT 16-APR-2005
DEFINITION	Streptococcus agalactiae NEM316 complete genome, segment 5.		
ACCESSION	AL766847	AL732656	
VERSION	AL766847.1	GI:23095200	
KEYWORDS	Streptococcus agalactiae NEM316		
SOURCE	Streptococcus agalactiae NEM316		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1		
AUTHORS	Glaser, P., Rueniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P. and Kunst, F.		
TITLE	Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease		
JOURNAL	Mol. Microbiol. 45 (6), 1499-1513 (2002)		
PUBMED	12354221		
REFERENCE	2		
AUTHORS	Glaser, P., Rueniok, C. and Frangeul, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86		
FEATURES	Location/Qualifiers		
source	1..167050		
	/organism="Streptococcus agalactiae NEM316"		
	/mol_type="genomic DNA"		
	/strain="NEM316"		
	/serovar="serovar III"		
	/db_xref="taxon:211110"		
gene	join(288..293,303..1412)		
RBS	288..293		
	/gene="gbs0767"		
	/gene="gbs0768"		
CDS	303..1412		
	/gene="gbs0769"		
	/note="Similar to riboflavin specific deaminase (diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil		
reductase)"			
/codon_start=1			
/transl_table=11			
/product="Unknown"			
/protein_id="CAD46411.1"			
/db_xref="GI:23095201"			
/db_xref="GOA:Q8E660"			
/db_xref="InterPro:IPR002125"			
/db_xref="InterPro:IPR002734"			
/db_xref="InterPro:IPR004794"			
/db_xref="InterPro:IPR011549"			
/db_xref="Uniprot/TrEMBL:Q8E660"			
/translation="MMKKEDYMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGVSL			
KRFCDLHAERQAIKNADEDISGSTLYVTLEPCCHVGKQPCTETALIKSGIKVVGSL			
DPNPLVSGKGIALLRKECLNVEGILREEDCALNERFIHMYTKQPFVYLKYMTLDG			
KIATKTGDSKWIISNEHSQVQLRQKCSAIMVGINTVLADNPRLTICRIKPGELVRI			
VCDSQLPLDSYLVKSAKTIPTWIATCSDNLAQQQTLKEMGRLIKVPRKGKLDLK			
VLMTILQEGIDSLILEGSSLHFSALKAGIVNRLIVFIAPKIIGGLAKTAISGEGL			
DWLNQDPVRVDIELSRMDSVDWIEGKVEHYVYRN"			
1381..1387			
/gene="gbs0768"			
1381..1387			
/gene="gbs0768"			
1393..2043			
/gene="gbs0768"			
1393..2043			
/gene="gbs0768"			
/note="Similar to ribiflavin synthase alpha chain"			
/codon_start=1			
/transl_table=11			
/product="Unknown"			
/protein_id="CAD46412.1"			
/db_xref="GI:23095202"			
/db_xref="Uniprot/TrEMBL:Q8E659"			
/translation="MFTGIIEGMQVSRIRNGIKSQQLSIDAPKLVPLLRKGDSVAVN			
GVCLTVLDKSETAFIADVMPESMWRSLAALRLHLSKVNLELALRSDRLRGHFLVGH			
DGVCKIEKIOKDDIAVRFSIDAPPSIMSYIIIEKGSVALDGLSLTVSFTSHSPRSVI			
PHYMAQTNLSLKKVGDLLNIEVDVLGKYAEKFLAPNTRNTHTSVMDWSFLSENGY"			
2046..3254			
/gene="gbs0769"			
2046..2051			
/gene="gbs0769"			
2061..3254			
/gene="gbs0769"			
/note="Similar to GTP cyclohydrolase II/			
3,4-dihydroxy-2-butanone-4-phosphate synthase"			
/codon_start=1			
/transl_table=11			
/product="Unknown"			
/protein_id="CAD46413.1"			
/db_xref="GI:23095203"			
/db_xref="GOA:Q8E658"			
/db_xref="InterPro:IPR000422"			
/db_xref="InterPro:IPR000926"			
/db_xref="Uniprot/TrEMBL:Q8E658"			
/translation="MAFSPKIKLLQDIKSGKVVVLMDDENRENEGDLICAAEMVTKES			
INFMAKPGKGLICLPUSNYAEKLELAQMAASHNTHETAFTSIDHLSTSTGISASD			
RALTAKRVNDSKAKDFRPGLHFLPLAKEGVLARNGHTATVDRICLRLGKKECGL			
CCETIMADGDSMKRDELLAPAKHDLAIATIKQLQDYRQREGGVVREIQLPTQFG			
HFTAYGVSEVVANKEHALVKGDISGDEVLCRLHSECLTGDFVHSLRCDCEQLANA			
LQOIEAEGRCVLLYMRQEGIGLINKLKAYHLQEEGLDLEANLALGFEQEDRDYGV			
SAQLLKDLGINSINLLTNPNPKIQQEAEGICVKNRVPLQVAATAYDLNLYLTKYKEM			
GHLLD"			
3255..3739			
/gene="gbs0770"			
3255..3260			
/gene="gbs0770"			
3269..3739			
/gene="gbs0770"			
/note="Similar to riboflavin synthase complex beta chain"			
/codon_start=1			
/transl_table=11			

```
/db_xref="UniProt/TREMBL:Q8E654"
/translation="MKFYLVRHGKQTNWLNREGRFGANGDSPLLEAEATEEELEEGQYLS
SIHFDVYSSDLGRADTVNILDANSCKPEIHVTPOLREWALGTLGCKIATMQAIY
PROMTAFYONPLOPKHDMFGAESLYOTHRVSEPLRSLSKNDVKVLI VGHGANLTAS
IRSLGQYQSLHYKDKLDNASLTIETHDFKDFNCULTWNDKSVLRQEVKMT"
complement(7064. .7069)
RBS
/gene="gbs0773"
complement(7080. .7553)
/gene="gbs0774"
complement(7080. .7553)
CDS
/gene="gbs0774"
/note="Similar to transcriptional regulator"
/codon_start=1
/transl_table=11

Query Match 100.0%; Score 23; DB 1; Length 167050;
Best Local Similarity 100.0%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0;

Qy 1 TGAGAAGCGAGTACAGAAAGCTTAG 23
Db 79838 TGAGAAGCGAGTACAGAAAGCTTAG 79860

RESULT 7
CO655069 349980 bp DNA linear PAT 02-FEB-2004
LOCUS
DEFINITION Sequence 12026 from Patent WO0234771.
ACCESSION CO655069
VERSION CO655069.1 GI:41687946
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
1 Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Frazer,C.
and Tettelin,H.
Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12026 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1. 349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/note="seq 10967 to big: splitted in
seq 10.967, from 0.000.001 to 0.349.980
seq 12.025, from 0.300.001 to 0.649.980
seq 12.026, from 0.600.001 to 0.949.980
seq 12.027, from 0.900.001 to 1.249.980
seq 12.028, from 1.200.001 to 1.549.980
seq 12.029, from 1.500.001 to 1.849.980
seq 12.030, from 1.800.001 to 2.149.980
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 100.0%; Score 23; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAGAAGCGAGTACAGAAAGCTTAG 23
Db 191615 TGAGAAGCGAGTACAGAAAGCTTAG 191637

RESULT 8
AX954529 349980 bp DNA linear PAT 08-JAN-2004
LOCUS
DEFINITION Sequence 1375 from Patent WO03093306.
ACCESSION AX954529
VERSION AX954529.1 GI:40783902
KEYWORDS

/product="Unknown"
/protein_id="CAD46414.1"
/db_xref="GI:23095204"
/db_xref="GOA:Q8E657"
/db_xref="UniProt:IPR002180"
/db_xref="UniProt/Swiss-Prot:Q8E657"
/translation="WTIEGOLVANEMKGIWVSRFNEILTSKLLSGAVDGLLRHGVG
BEDIDIVWPGAFIIPYMARWALYKDYDAICIGVVIKGSTHDYDVYCNVETKIGH
LNSQSDPHIFGVITDNEQAIERAGTKAGNKGYDCALSAIEMVNLDKLRLR"
complement(3814. .5315)
gene
/gene="lyss"
complement(3814. .5304)
CDS
/gene="lyss"
/codon_start=1
/transl_table=11
/product="lyssyl-tRNA synthetase"
/protein_id="CAD46415.1"
/db_xref="GI:23095205"
/db_xref="GOA:Q8E656"
/db_xref="InterPro:IPR002313"
/db_xref="InterPro:IPR004364"
/db_xref="InterPro:IPR004365"
/db_xref="InterPro:IPR006195"
/db_xref="InterPro:IPR008994"
/db_xref="UniProt/Swiss-Prot:Q8E656"
/translation="MSNOHIELNDQIIVREKMAALTEQIGIDPGKFEFRTATSGOL
NEKYADSKEDLHDEETATTAGRLMTGRGKGVGFPAHQDREGQIQIYVRKDSVGE
NYEIPKADLGDGVEQVMTDMGSLSIKATHILSKALRPLPEKPHGLTDIETI
YRKRLHDIISNRDSDRVTRTKIIEIRREMDNSNGFLEVTVPVHLNAGSARPF
THNAQDIDMVLRTATELHLKRLIVGGMRVVEIGRIFRNEGMATHTNPFTSIEAYQ
AYADQDMDLATEGIIOHVTKVKGDPINYGQTEIKINEPKRVHMDVAKEITGID
FWKNTLSEAAQALQKRVPLEKHPTTYTGHLLINAFEFSEFVETLIQPTVFVGHVPEVS
PLAKNDDPDLRFELFIMTKYANATFELNDPDLQSLRFEQAQSAKELQDDEATG
VDYDYVEALEYGMPTGGIGIDRLCLMLLDTTITIRVDLLVFPPTMK"
complement(5310. .5315)
RBS
/gene="lyss"
5467. .6411
5467. .6411
/gene="gbs0772"
5467. .5472
/gene="gbs0772"
5479. .6381
/gene="gbs0772"
/note="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46416.1"
/db_xref="GI:23095206"
/db_xref="GOA:Q8E655"
/db_xref="InterPro:IPR005833"
/db_xref="InterPro:IPR005834"
/db_xref="InterPro:IPR006439"
/db_xref="UniProt/TREMBL:Q8E655"
/translation="MTISIVFDVDTIYDQQAQRIAYVEKCPDFDMSAINQAIYRFR
HVSIDIGFPPVWAGETVEYFRWCKETLLEFGYEIDEATGIYQEIVHELENITM
LDERMTDLFLKSKNVPVGMIIITNGTEHQLKVKVKGILYDVPDPRVLSQATQKPK
EKEIFNLAAEQDMPSTIYVGDSDINDIMAFNGWHSMFNRHRSLSKPGKIPVY
DVAINDFQLGAVKVLFDLPNKFIFDINDKSNPVLNGLNGLMMAERLLESNMS
VQKVVILLRTAKQKVLRMKYAR"
6390. .6411
terminator
/gene="gbs0772"
complement(6417. .7069)
/gene="gbs0773"
complement(6417. .7058)
CDS
/gene="gbs0773"
/note="Similar to putative phosphoglycerate mutase"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46417.1"
/db_xref="GI:23095207"
/db_xref="GOA:Q8E654"
/db_xref="InterPro:IPR001345"
```

```

SOURCE      Streptococcus pyogenes
ORGANISM    Streptococcus pyogenes
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.

REFERENCE   1
AUTHORS     Telford, J., Masignani, V., margarit y Ros, I., Grandi, G., Fraser, C.
            and Tettelin, H.
TITLE       Nucleic acids and proteins from streptococcus groups a b
JOURNAL     Patent: WO 03093306-A 1375 13-NOV-2003;
            Chiron SRL (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES    Location/Qualifiers
            1..349980
               /organism="Streptococcus pyogenes"
               /mol_type="unassigned DNA"
               /db_xref="taxon:1314"
               /note="seq 1373 to long, original length: 2.160.266
            replaced by
            seq 1373: from 0.000.001 to 0.349.980
            seq 1374: from 0.300.001 to 0.649.980
            seq 1375: from 0.600.001 to 0.949.980
            seq 1376: from 0.900.001 to 1.249.980
            seq 1377: from 1.200.001 to 1.549.980
            seq 1378: from 1.500.001 to 1.849.980
            seq 1379: from 1.800.001 to 2.149.980
            seq 1380: from 2.100.001 to 2.160.266"

ORIGIN
Query Match      100.0%; Score 23; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 220402 TGAGAGGCAGTAGAAGCTTAG 220424

RESULT 9
AC123147
LOCUS      AC123147          185291 bp    DNA    linear    HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-333E19, *** SEQUENCING IN PROGRESS
            ***
ACCESSION  AC123147.4   GI:24942304
VERSION    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
            1 (bases 1 to 185291)
Munzy,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Centek,A.,
Chacko,J., Chavez,D., Chen,G., Chen,K., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,R., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

```

```

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemeleleh,O., Okwunu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojao,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaaj,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaeana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 185291)
Worley,K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 185291)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23603979.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMGN
Center clone name: CH230-333E19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 172302 bases at least Q40
Consensus quality: 174607 bases at least Q30
Consensus quality: 175894 bases at least Q20
Estimated insert size: 181132; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs

```

\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 185291: contig of 185291 bp in length.

## FEATURES

source  
 1..185291  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-333B19"

## misc\_feature

1..1735  
 /note="wgs\_contig"

## misc\_feature

178867..179717  
 /note="clone\_boundary"

clone\_end:77

## misc\_feature

end\_sequence:B2234987"  
 183599..185291  
 /note="wgs\_end\_extension"  
 clone\_end:77"

## ORIGIN

Query Match 91.3%; Score 21; DB 14; Length 185291;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAAAGCTTA 22

Db 59437 GAGAGGCGAGTAAAGCTTA 59457  
 |||||  
 |||||

## RESULT 10

AC121422/c

## LOCUS

AC121422 231447 bp DNA linear HTG 15-NOV-2002  
 Rattus norvegicus clone CH230-11588, WORKING DRAFT SEQUENCE, 2  
 unordered pieces.

## DEFINITION

AC121422 4 GI:25008025

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

## KEYWORDS

Rattus norvegicus (Norway rat)

## SOURCE

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

## AUTHORS

1 (bases 1 to 231447)  
 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensuewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Nwakelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parke, K.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rui, S. J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaje, D.,  
 Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstock, G. and Gibbs, R. A.

## Direct Submission

Unpublished

2 (bases 1 to 231447)

Worley, K. C.

## Direct Submission

Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231447)

## Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Nov 15, 2002 this sequence version replaced gi:23196017.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

## Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXIS

Center clone name: CH230-11588

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 211543 bases at least Q40

Consensus quality: 213418 bases at least Q30

Consensus quality: 214985 bases at least Q20

Estimated insert size: 215208; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\*

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 227883: contig of 227883 bp in length
* 227884 227983: gap of unknown length
* 227984 231447: contig of 3464 bp in length.

FEATURES             source
    Location/Qualifiers
        1..231447
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-115B8"
    misc_feature
        1..1431
        /note="wgs_end_extension
        clone_end:Sp6"
    misc_feature
        2174..2753
        /note="clone_boundary
        clone_end:Sp6
        site:
        end_sequence:BH292203"
    misc_feature
        complement(221179..221702)
        /note="clone_boundary
        clone_end:T7
        site:
        end_sequence:BH292202"
    misc_feature
        224336..225446
        /note="wgs_end_extension
        clone_end:T7"
    misc_feature
        226785..227883
        /note="wgs_end_extension
        clone_end:T7"
    gap
        227884..227983
        /estimated_length=unknown
    misc_feature
        227984..229146
        /note="wgs_end_extension
        clone_end:T7"
    misc_feature
        229860..231447
        /note="wgs_end_extension
        clone_end:T7"

ORIGIN
Query Match          91.3%; Score 21; DB 14; Length 231447;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GAGAGGCAGTAGAAGCTTA 22
|||||
Db 183282 GAGAGGCAGTAGAAGCTTA 183262

RESULT 11
AC095375
LOCUS          AC095375          306616 bp          DNA          linear          HTG 13-NOV-2002
DEFINITION    Rattus norvegicus clone CH230-2E1, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION    AC095375
VERSION      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 306616)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alabrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

```

```

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Loulseged, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uemami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 306616)
Worley, K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 306616)
Worley, K.C.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23264771.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
----- Project name: TUOF
Center clone name: CH230-2E1
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 244944 bases at least Q40
Consensus quality: 247400 bases at least Q30
Consensus quality: 249174 bases at least Q20

```

Estimated insert size: 250113; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/GenBank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 260565: contig of 260565 bp in length  
\* 260566: gap of unknown length  
\* 301082: contig of 40417 bp in length  
\* 301083 301182: gap of unknown length  
\* 301183 302346: contig of 1164 bp in length  
\* 302347 302446: gap of unknown length  
\* 302447 306616: contig of 4170 bp in length.

## FEATURES

source  
1..306616

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-2E1"

misc\_feature

1..1067  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature

4919..6420  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature

5053..5785  
/note="clone\_boundary  
clone\_end:T7  
site\_

misc\_feature

end\_sequence:BH262876"

6620..7177  
/note="clone\_boundary  
clone\_end:Sp6  
site\_

gap

end\_sequence:BH262877"

misc\_feature

260566..260665  
/estimated\_length=unknown  
260666..261775  
/note="wgs\_end\_extension  
clone\_end:Sp6"

misc\_feature

26331..264680  
/note="wgs\_end\_extension  
clone\_end:Sp6"

gap

301083..301182

gap

/estimated\_length=unknown  
302347..302446  
/estimated\_length=unknown

## ORIGIN

Query Match 88.7%; Score 20.4; DB 14; Length 306616;  
Best Local Similarity 95.5%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGCTTAG 23  
|||||

DB 160683 GAGAGGCAGTAGAAGCTTAG 160704  
|||||

## RESULT 12

AR353664 AR353664 1378 bp DNA linear PAT 17-AUG-2003  
LOCUS  
DEFINITION Sequence 14 from patent US 6593093.  
ACCESSION AR353664  
VERSION AR353664.1 GI:33759697  
KEYWORDS

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1378)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 14 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1..1378  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 86.1%; Score 19.8; DB 6; Length 1378;  
Best Local Similarity 91.3%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 9 TGAATATGCAGTAGAAGCTTAG 31  
|||||  
RESULT 13  
AX817901 AX817901 1378 bp DNA linear PAT 10-DEC-2003  
LOCUS  
DEFINITION Sequence 14 from Patent EP1338656.  
ACCESSION AX817901  
VERSION AX817901.1 GI:39723087  
KEYWORDS  
SOURCE Streptococcus sp. 'group A'  
ORGANISM Streptococcus sp. 'group A'  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Uhl,J.R. and Cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 14 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
1..1378  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptai sequence from isolate no. 10"  
ORIGIN  
Query Match 86.1%; Score 19.8; DB 6; Length 1378;  
Best Local Similarity 91.3%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 9 TGAATATGCAGTAGAAGCTTAG 31  
|||||  
RESULT 14  
AR353657 AR353657 1379 bp DNA linear PAT 17-AUG-2003  
LOCUS  
DEFINITION Sequence 7 from patent US 6593093.  
ACCESSION AR353657  
VERSION AR353657.1 GI:33759690  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1379)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 7 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1..1379

ORIGIN	/organism="unknown"		/mol_type="genomic DNA"		
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2; Indels 0; Gaps 0;
Qy	1 TGAGAAGGCAGTAGAAGCTTAG 23				
Db	9 TGAATAATGCAGTAGAAGCTTAG 31				
RESULT 15					
AX817894	1379 bp DNA linear PAT 17-AUG-2003				
LOCUS					
DEFINITION	Sequence 13 from patent US 6593093.				
ACCESSION	AR353663				
VERSION	AR353663.1 GI:33759666				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1379)				
AUTHORS	Uhl,J.R. and Cockerill Iii,F.R.				
TITLE	Detection of group a Streptococcus				
JOURNAL	Patent: US 6593093-A 13 15-JUL-2003;				
Mayo Foundation for Medical Education and Research; Rochester, MN					
FEATURES	Location/Qualifiers				
source	1..1379				
/organism="unknown"					
/mol_type="genomic DNA"					
ORIGIN					
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2; Indels 0; Gaps 0;
Qy	1 TGAGAAGGCAGTAGAAGCTTAG 23				
Db	9 TGAATAATGCAGTAGAAGCTTAG 31				
RESULT 16					
AX817894	1379 bp DNA linear PAT 17-AUG-2003				
LOCUS					
DEFINITION	Sequence 16 from patent US 6593093.				
ACCESSION	AR353666				
VERSION	AR353666.1 GI:33759699				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1379)				
AUTHORS	Uhl,J.R. and Cockerill Iii,F.R.				
TITLE	Detection of group a Streptococcus				
JOURNAL	Patent: US 6593093-A 15 15-JUL-2003;				
Mayo Foundation for Medical Education and Research; Rochester, MN					
FEATURES	Location/Qualifiers				
source	1..1379				
/organism="unknown"					
/mol_type="genomic DNA"					
ORIGIN					
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2; Indels 0; Gaps 0;
Qy	1 TGAGAAGGCAGTAGAAGCTTAG 23				
Db	10 TGAATAATGCAGTAGAAGCTTAG 32				
RESULT 17					
AX817894	1379 bp DNA linear PAT 10-DEC-2003				
LOCUS					
DEFINITION	Sequence 7 from Patent EP1338656.				
ACCESSION	AX817894				
VERSION	AX817894.1 GI:39723080				
KEYWORDS	Streptococcus sp. 'group A'				
SOURCE	Streptococcus sp. 'group A'				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
REFERENCE	1				
AUTHORS	Uhl,J.R. and Cockerill Iii,F.R.				
TITLE	Detection of group a Streptococcus				
JOURNAL	Patent: EP 1338656-A 7 27-AUG-2003;				
Mayo Foundation for Medical Education and Research (US)					
FEATURES	Location/Qualifiers				
source	1..1379				
/organism="Streptococcus sp. 'group A'"					
/mol_type="unassigned DNA"					
/db_xref="taxon:36470"					
/note="ptSI sequence from isolate no. 6"					
ORIGIN					
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2; Indels 0; Gaps 0;
Qy	1 TGAGAAGGCAGTAGAAGCTTAG 23				
Db	9 TGAATAATGCAGTAGAAGCTTAG 31				
RESULT 18					
AX817900	1379 bp DNA linear PAT 10-DEC-2003				
LOCUS					
DEFINITION	Sequence 13 from Patent EP1338656.				
ACCESSION	AX817900				
VERSION	AX817900.1 GI:39723086				
KEYWORDS	Streptococcus sp. 'group A'				
SOURCE	Streptococcus sp. 'group A'				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
REFERENCE	1				
AUTHORS	Uhl,J.R. and Cockerill Iii,F.R.				
TITLE	Detection of group a Streptococcus				
JOURNAL	Patent: EP 1338656-A 13 27-AUG-2003;				
Mayo Foundation for Medical Education and Research (US)					
FEATURES	Location/Qualifiers				
source	1..1379				
/organism="Streptococcus sp. 'group A'"					
/mol_type="unassigned DNA"					
/db_xref="taxon:36470"					
/note="ptSI sequence from isolate no. 4"					
ORIGIN					
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2; Indels 0; Gaps 0;
Qy	1 TGAGAAGGCAGTAGAAGCTTAG 23				
Db	9 TGAATAATGCAGTAGAAGCTTAG 31				
RESULT 19					
AX817903	1379 bp DNA linear PAT 10-DEC-2003				
LOCUS					
DEFINITION	Sequence 16 from Patent EP1338656.				
ACCESSION	AX817903				
VERSION	AX817903.1 GI:39723089				
KEYWORDS					



Query Match	86.1%	Score 19.8	DB 6	Length 1384
Best Local Similarity	91.3%	Pred. No. 1.2e+03		
Matches	0	Mismatches	2	Indels 0
Match 21	Conservative			Gaps 0

```
Qy 1 TGAGAGGCAGTAGAAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAAGCTTAG 31

RESULT 24
AR353659
LOCUS AR353659 1385 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 9 from patent US 6593093.
ACCESSION AR353659
VERSION AR353659.1 GI:33759692
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1385)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group a Streptococcus
JOURNAL Patent: US 6593093-A 9 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
FEATURES
source
1..1385
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1385;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAAGCTTAG 32

RESULT 25
AR353659
LOCUS AR353659 1385 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 9 from Patent EP1338656.
ACCESSION AR353659
VERSION AR353659.1 GI:39723082
KEYWORDS
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus sp. 'group A'
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE
1
AUTHORS Uhl,J.R. and Cockerill Iii,F.R.
TITLE Detection of group a Streptococcus
JOURNAL Patent: EP 1338656-A 9 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
source
1..1385
/organism="Streptococcus sp. 'group A'"
/db_xref="taxon:36470"
/note="ptsi sequence from isolate no. 7"

ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1385;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAAGCTTAG 32

RESULT 26
AR353661
LOCUS AR353661 1390 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6593093.
ACCESSION AR353661
VERSION AR353661.1 GI:33759694
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1390)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group a Streptococcus
JOURNAL Patent: US 6593093-A 11 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
FEATURES
source
1..1390
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAAGCTTAG 23
Db 24 TGAATAATGCAGTAGAAAGCTTAG 46

RESULT 27
AR353665
LOCUS AR353665 1393 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 11 from Patent EP1338656.
ACCESSION AR353665
VERSION AR353665.1 GI:39723084
KEYWORDS
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus sp. 'group A'
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE
1
AUTHORS Uhl,J.R. and Cockerill Iii,F.R.
TITLE Detection of group a Streptococcus
JOURNAL Patent: EP 1338656-A 11 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
source
1..1390
/organism="Streptococcus sp. 'group A'"
/db_xref="taxon:36470"
/note="ptsi sequence from isolate no. 9"

ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAAGCTTAG 23
Db 24 TGAATAATGCAGTAGAAAGCTTAG 46

RESULT 28
AR353665
LOCUS AR353665 1393 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 15 from patent US 6593093.
ACCESSION AR353665
VERSION AR353665.1 GI:33759698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1393)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group a Streptococcus
JOURNAL Patent: US 6593093-A 15 15-JUL-2003;
```

ORIGIN	
Query Match	86.1%; Score 19.8; DB 6; Length 1447;
Best Local Similarity	91.3%; Pred. NO. 1.2e+02;
Matches 21: Conservative	0; Mismatches 2; Indels 0; Gaps 0;

AR353656	AR353656	1803 bp	DNA	linear	PAT 17-AUG-2003
LOCUS	Sequence 5 from patent US 6593093.				
DEFINITION	AR353656				
ACCESSION	AR353656				
VERSION	AR353656.1	GI:33759689			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1803)				
AUTHORS	Uhl, J.R. and Cockerill, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: US 6593093-A 5 15-JUL-2003;				
JOURNAL	Mayo Foundation for Medical Education and Research; Rochester, MN				
FEATURES	Location/Qualifiers				
source	1..1803				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	86.1%; Score 19.8; DB 6; Length 1803;				
Best Local Similarity	91.3%; Pred. No. 1.2e+02;				
Matches	21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1 TGAGAGGCAGTAGAAGCTTAG 23				
Db					
177 TGAAGATCGAGTAGAAGCTTAG 199					
RESULT 34					
LOCUS	AX817892	1803 bp	DNA	linear	PAT 10-DEC-2003
DEFINITION	Sequence 5 from Patent EP1338656.				
ACCESSION	AX817892				
VERSION	AX817892.1	GI:39723079			
KEYWORDS	Streptococcus sp. 'group A'				
SOURCE	Streptococcus sp. 'group A'				
ORGANISM	Streptococcus sp. 'group A'				
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
	Streptococcus.				
REFERENCE	1				
AUTHORS	Uhl, J.R. and Cockerill, Iii, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: EP 1338656-A 5 27-AUG-2003;				
JOURNAL	MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)				
FEATURES	Location/Qualifiers				
source	1..1803				
	/organism="Streptococcus sp. 'group A'"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:36470"				
	/note="ptsl sequence from Oklahoma University M1 strain"				
ORIGIN					
Query Match	86.1%; Score 19.8; DB 6; Length 1803;				
Best Local Similarity	91.3%; Pred. No. 1.2e+02;				
Matches	21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1 TGAGAGGCAGTAGAAGCTTAG 23				
Db					
177 TGAAGATCGAGTAGAAGCTTAG 199					
RESULT 35					
LOCUS	AE006575/c	10430 bp	DNA	linear	BCT 03-JUN-2004
DEFINITION	Streptococcus pyogenes M1 GAS, section 104 of 167 of the complete genome.				
ACCESSION	AE006575	AE004092			
VERSION	AE006575.1	GI:13622480			
KEYWORDS	Streptococcus pyogenes M1 GAS				
SOURCE	Streptococcus pyogenes M1 GAS				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
REFERENCE	1 (bases 1 to 10430)				
AUTHORS	Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.				
TITLE	Complete genome sequence of an M1 strain of Streptococcus pyogenes				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)				
JOURNAL	11296296				
REFERENCE	2 (bases 1 to 10430)				
AUTHORS	Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104, USA				
FEATURES	Location/Qualifiers				
source	1..10430				
	/organism="Streptococcus pyogenes M1 GAS"				
	/mol_type="genomic DNA"				
	/strain="SF370"				
	/serotype="M1"				
	/db_xref="taxon:160490"				
	complement(78..1505)				
gene	/gene="gapN"				
	/locus_tag="SPy1371"				
CDS	complement(78..1505)				
	/gene="gapN"				
	/locus_tag="SPy1371"				
	/EC_number="1.2.1.9"				
	/note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of S. pyogenes. Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of Streptococcus pyogenes proteins. Unpublished data.; Best Blastp hit = dbj BAB05956.1  (AP001514) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase [Bacillus halodurans]"				
	/codon_start=1				
	/transl_table=11				
	/product="putative NADP-dependent				
	glyceraldehyde-3-phosphate dehydrogenase"				
	/protein_id="AAK34198.1"				
	/db_xref="GI:13622481"				
	/translation="MAKQYKNLVNGEWKLSENEITTYAPATGSELGSPVPMTOAEVDA VYASAKKALSDWRALSVERAAVLHKAADILVRDAEKIGAILSKVEAKGHKAAVSEVI RTAEIINAAEEGLRMEGEVLGGSPFAASKKIAIVRREPVGIVLAISPNYVNVLA GSKIAFALTAGNVVVALKPPPTQGSISGLLLAEFAEAGIPAGVNTITGRSGVIGDIY EHSVSPFINFTGTPIGEGIGKLGMRPIMLELGGKDSALVLEDADJALAAKNIVAGA EGYSGQRCTAVKRVLVMDKVDQDLAAEIKTLVLEKLSVGMPEDDADITPLIDTSAADFV EGLIKDATDKGATALTAFNREGNLISPLVDHVTMDRLAWEEPFGVLPFIIRVTVE EALKISNESYGLQASIFTTNFPKAFGIAEQLEVTGTVHLNNKTQRTGDTNFPFLGAKKS GAGVQGVKYIEAMTTVKSVVDIQ"				
	complement(1690..3423)				
gene	/gene="pstI"				
	/locus_tag="SPy1372"				
CDS	complement(1690..3423)				
	/gene="pstI"				
	/locus_tag="SPy1372"				
	/EC_number="2.7.3.9"				
	/note="Best Blastp hit = sp P30299 PT1 STRSL PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME 1) >gi 322087 pir  JCI1375 phosphotransferase system enzyme I (EC 2.7.3.9) - Streptococcus salivarius >gi 153615 gb AA26873.1  (M81756) phosphoenolpyruvate:sugar phosphotransferase system enzyme I [Streptococcus salivarius] >gi 1583334 prf  2120381B ptsl gene [Streptococcus salivarius]"				
	/codon_start=1				



group A Streptococcus strains associated with acute rheumatic fever outbreaks  
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
11917108  
2 (bases 1 to 11992)  
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Musser,J.M.  
Direct Submission  
Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St., Hamilton, MT 59840, USA

FEATURES

source

1..11992 Location/Qualifiers  
/organism="Streptococcus pyogenes MGAS8232"  
/mol\_type="genomic DNA"  
/strain="MGAS8232"  
/db\_xref="taxon:186103"  
132..1073  
/gene="bira"  
/note="synonym: spym18\_1374"  
132..1073  
/gene="bira"  
/note="best blastp match gb|AAK34189.1| (AE006574)  
putative biotin operon repressor [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAL97969.1"  
/db\_xref="GI:19748528"  
/translation="MKVSEKYYQLLSQTDVSGEYLADQLISRTSVKSIKLENQ GIQDLSLKHGRVQGDILLPKTIISQGLGMPVITYPHSOSTQDADQGEAHNSAPR LYLAPQEAQRGRDFFSASTGCIYMSYLNKPNVPYAMPPTMVAASSIVKAISR LTGDTETIWNVDIYGNHKGAVILTAITSVETGLITDVIIGVGLNFFVTDPEAIA QKAGSLFTEKPTITRNDLIIDIWKLFLSIPVKDHVKVYKESLVNLNQVTFIENSQEK RAIADLTQDGLHVQPDNGDQLTIRSGEISLSSWEN"  
complement (1048..1320)  
/gene="spym18\_1375"  
complement (1048..1320)  
/gene="spym18\_1375"  
/note="best blastp match gb|AAK34190.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97970.1"  
/db\_xref="GI:19748529"  
/translation="MPLPTFYLVTFEHSTVFECYNKAWTIRQFLFMAFVCAFEITYFEN DLLLSGNLYLFAFWGLLFRDLRRVHTINQLTKILKAANSFKKXD"  
complement (1346..3016)  
/gene="dnaX"  
/note="synonym: spym18\_1376"  
complement (1346..3016)  
/gene="dnaX"  
/note="best blastp match gb|AAF98348.1| (AF280764) DNA polymerase III gamma/tau subunits [Streptococcus pyogenes]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAL97971.1"  
/db\_xref="GI:19748530"  
/translation="MYQALYRKYSOTPFDEMVGQSVISTTLKQAVESGKISHAYLFSG PRGTGKTSAAKIFAKAMNCPQVQGPCNQCDICRIDTNGSLDEVIDAASNGVDE IRDRLDSTYAPSRATYKVYIIDVHMLSTGAFNALLKTLEPTENVVFILATLELKH IPATLIRSVORFEFKAIKQKAIREHLAWLDKEGIAVEVDALNLIARRAGGMDALS ILDQALSLSPNQVITIAEISITGISILALGCVRYVSQEQATQALAALETIYDSGK SMSRFATLLYLRDLVVKAGGNQKQSAVFDNLSLIDRIQFMITTVTSQLOPEIK KGTPRIYAENMTIQAQKEQILSQANLSGLIEIETLKNELQAKQKQSQLOSRPD SLARSDTKPTSTGYRVDRTVILKMEETVNSQSQRYDLDALKNWNEILDNISAQD RALLWGSEPVLANSENAILAPEAAFNAEQVMNSRNNDMFGNIMSKAAGFSFNILAVP

gene

CDS

RTDFQHIRKEFAQMKSQKDSVQEQEVALDIPGEFDFLLDKINTIDD"  
complement (3016..3513)  
/gene="spym18\_1377"  
complement (3016..3513)  
/gene="spym18\_1377"  
/note="best blastp match gb|AAK34192.1| (AE006574)  
conserved hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAL97972.1"  
/db\_xref="GI:19748531"  
/translation="MNKSKIEOYQOLMIAQAKELFANESNALNLSASALLNWTLEN SVFTGPLYDGOELILGPFQGRVSCVHIKLGKVGCSASQSRITLINDVKQAHNYIS CDAAMSEIVVPMVKEGHLIGVLDSSLVADYDEVQEYLETFFVLLLEKTYTFNM FGVKN"  
3658..4467  
/gene="spym18\_1378"  
3658..4467  
/gene="spym18\_1378"  
/note="best blastp match gb|AAK34193.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97973.1"  
/db\_xref="GI:19748532"  
/translation="MKQITQCKWLRYGLFIALVNLGIELLELGLTANNLSFKKFAFALI LTIISLGIYLIFFAAAI FYLSCKFHMNLNVIIVSCISGLYISGLASCGNHLVGQFWS YIIPSKDALKMGDALTAPIVEEPIKASSAILVITLPPRLTLKELVALLSGMGFOL TEDIRYLQAKSIDSLVPTAIGRISTAVTSHWVHTAIFTIGAYLLKGSNLSKQKQOI FWLLSPLVHFIWNPSPLTSIPGNTVLLGTLLILFGDLFQKINTLDDVLF"  
complement (4520..4783)  
/gene="spym18\_1379"  
complement (4520..4783)  
/gene="spym18\_1379"  
/note="best blastp match gb|AAK34194.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97974.1"  
/db\_xref="GI:19748533"  
/translation="MLLIWGLLFVWNELVFIYAPDKRAIKKGRISERNLLVITV LFGFGALLAAKXHHKTRKWFYITWYTSILTLVLYIFMK"  
complement (4863..5489)  
/gene="udk"  
/note="synonym: spym18\_1380"  
complement (4863..5489)  
/gene="udk"  
/note="best blastp match gb|AAK34195.1| (AE006574)  
putative uridine kinase [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="putative uridine kinase"  
/protein\_id="AAL97975.1"  
/db\_xref="GI:19748534"  
/translation="MFKKPIIIGVTGSGGKTSVSRAILDSPNARIAMIQHDSYRK DQSHMSFEERVKNYDHPADFMIQQLKELLAGRPVDPIYDYKXGTRSNTRFQ DPQDVIIEGILVLEDERLMDIKLFDVTDKDDIIRIRIKRDMMERGRSLESIIQD YTSVVRPMYHQFIEPSKRYADIVIPEGVSNVVALDIVNSKIASILGEV"  
5587..6672  
/gene="spym18\_1381"  
5587..6672  
/gene="spym18\_1381"  
/note="best blastp match gb|AAK34196.1| (AE006574)  
putative RNA helicase [Streptococcus pyogenes M1 GAS]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="putative RNA helicase"  
/protein\_id="AAL97976.1"  
/db\_xref="GI:19748535"







Db 107626 TGAAAAATGCAGTAGAAGCTTAG 107648

RESULT 39  
BA000034\_08  
WPCOMMENT

Sequence split into 19 fragments LOCUS BA000034 Accession BA000034

Fragment Name	Begin	End
BA000034_00	1	110000
BA000034_01	100001	210000
BA000034_02	200001	310000
BA000034_03	300001	410000
BA000034_04	400001	510000
BA000034_05	500001	610000
BA000034_06	600001	710000
BA000034_07	700001	810000
BA000034_08	800001	910000
BA000034_09	900001	1010000
BA000034_10	1000001	1110000
BA000034_11	1100001	1210000
BA000034_12	1200001	1310000
BA000034_13	1300001	1410000
BA000034_14	1400001	1510000
BA000034_15	1500001	1610000
BA000034_16	1600001	1710000
BA000034_17	1700001	1810000
BA000034_18	1800001	1894275

Continuation (9 of 19) of BA000034 from base 800001 (BA000034 Streptococcus pyogenes SS1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 7626 TGAAAAATGCAGTAGAAGCTTAG 7648

RESULT 40  
CP000003\_10/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000003 Accession CP000003

Fragment Name	Begin	End
CP000003_00	1	110000
CP000003_01	100001	210000
CP000003_02	200001	310000
CP000003_03	300001	410000
CP000003_04	400001	510000
CP000003_05	500001	610000
CP000003_06	600001	710000
CP000003_07	700001	810000
CP000003_08	800001	910000
CP000003_09	900001	1010000
CP000003_10	1000001	1110000
CP000003_11	1100001	1210000
CP000003_12	1200001	1310000
CP000003_13	1300001	1410000
CP000003_14	1400001	1510000
CP000003_15	1500001	1610000
CP000003_16	1600001	1710000
CP000003_17	1700001	1810000
CP000003_18	1800001	1899877

Continuation (11 of 19) of CP000003 from base 1000001 (CP000003 Streptococcus pyogenes M

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 93181 TGAAAAATGCAGTAGAAGCTTAG 93159

RESULT 41  
CP0000017\_10/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP0000017 Accession CP0000017

Fragment Name	Begin	End
CP0000017_00	1	110000
CP0000017_01	100001	210000
CP0000017_02	200001	310000
CP0000017_03	300001	410000
CP0000017_04	400001	510000
CP0000017_05	500001	610000
CP0000017_06	600001	710000
CP0000017_07	700001	810000
CP0000017_08	800001	910000
CP0000017_09	900001	1010000
CP0000017_10	1000001	1110000
CP0000017_11	1100001	1210000
CP0000017_12	1200001	1310000
CP0000017_13	1300001	1410000
CP0000017_14	1400001	1510000
CP0000017_15	1500001	1610000
CP0000017_16	1600001	1710000
CP0000017_17	1700001	1810000
CP0000017_18	1800001	1838554

Continuation (11 of 19) of CP0000017 from base 1000001 (CP0000017 Streptococcus pyogenes 1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 97184 TGAAAAATGCAGTAGAAGCTTAG 97162

RESULT 42  
CP0000056\_11/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP0000056 Accession CP0000056

Fragment Name	Begin	End
CP0000056_00	1	110000
CP0000056_01	100001	210000
CP0000056_02	200001	310000
CP0000056_03	300001	410000
CP0000056_04	400001	510000
CP0000056_05	500001	610000
CP0000056_06	600001	710000
CP0000056_07	700001	810000
CP0000056_08	800001	910000
CP0000056_09	900001	1010000
CP0000056_10	1000001	1110000
CP0000056_11	1100001	1210000
CP0000056_12	1200001	1310000
CP0000056_13	1300001	1410000
CP0000056_14	1400001	1510000
CP0000056_15	1500001	1610000
CP0000056_16	1600001	1710000
CP0000056_17	1700001	1810000
CP0000056_18	1800001	1897573

Continuation (12 of 19) of CP0000056 from base 1100001 (CP0000056 Streptococcus pyogenes 1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 20489 TGAAAAATGCAGTAGAAGCTTAG 20467

RESULT 43  
AC133332  
LOCUS

AC133332 140072 bp DNA linear HTG 10-SEP-2002

```

DEFINITION Felis catus clone RP86-331P23, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC133332
VERSION AC133332.1 GI:22773276
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Felis catus (cat)
ORGANISM Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
Felinae; Felis.
1 (bases 1 to 140072)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Hegnighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaepi,R., Maduro,Q.B., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 140072)
Green,E.D.
Direct Submission
Submitted (10-SEP-2002) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoehngri.nih.gov
----- Project Information
Center project name: dao
Center clone name: 331P23
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 138836 bases at least Q40
Consensus quality: 138960 bases at least Q30
Consensus quality: 139106 bases at least Q20
Insert size: 111000; agarose-fp
Insert size: 139772; sum-of-contigs
Quality coverage: 12.82x in Q20 bases; agarose-fp
Quality coverage: 10.18x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2120: contig of 2120 bp in length
* 2121 2220: gap of unknown length
* 2221 33322: contig of 31102 bp in length
* 33323 33422: gap of unknown length
* 33423 77664: contig of 44242 bp in length
* 77665 140072: gap of unknown length
* 140072: contig of 62308 bp in length.
Location/Qualifiers
1. .140072
/organism="Felis catus"
/mol_type="genomic DNA"
/db_xref="taxon:9685"
/clone="RP86-331P23"
/clone_lib="RP86"
1. .2120
/note="assembly_fragment"
2121. .2220

FEATURES
source
1. .140072
/note="estimated length=unknown"
2221. .33322
/note="assembly_fragment
clone_end:SP6
vector_side:right"
33323. .33422
/estimated length=unknown
33423. .77664
/note="assembly_fragment
clone_end:T7
vector_side:right"
77665. .77764
/estimated length=unknown
77765. .140072
/note="assembly_fragment"

misc_feature
86.1%; Score 19.8; DB 14; Length 140072;
Best Local Similarity 91.3%; Pred. No. 82;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGCGCAGTAGAAAGCTTAG 23
| | | | | | | | | | | | | | | | | | | | |
Db 17702 TAAGAGGCTAGTAGAAGCTTAG 17724

RESULT 44
AC154824/c
LOCUS AC154824 164494 bp DNA linear ROD 21-JUN-2005
DEFINITION Mus musculus BAC clone RP24-540A17 from chromosome 13, complete
sequence.
ACCESSION AC154824
VERSION AC154824.2 GI:66773639
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 164494)
Adams,S. and Haglund,K.
TITLE The sequence of Mus musculus BAC clone RP24-540A17
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 164494)
AUTHORS Wilson,R.K.
TITLE Direct Submission
Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 164494)
AUTHORS Wilson,R.K.
TITLE Direct Submission
Submitted (14-JAN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 164494)
AUTHORS Wilson,R.K.
TITLE Direct Submission
Submitted (27-MAY-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 164494)
AUTHORS Wilson,R.K.
TITLE Direct Submission
Submitted (21-JUN-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 27, 2005 this sequence version replaced gi:56900450.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: M_BB0540A17
Drafting center: WIBR

```

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.

If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

```

FEATURES             Location/Qualifiers
     source            1..164494
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosomes="13"
                        /clone_lib="RPCI-24"
                        /clone="RP24-540A17"
                        /notes="Sequence derived from PCR product of project DNA"
     misc_feature      14473..14609
                        /notes="Sequence derived from PCR product of project DNA"
     misc_feature      16026..16246
                        /notes="Sequence derived from PCR product of project DNA"
     misc_feature      16620..16628
                        /notes="Sequence derived from PCR product of project DNA"
     unsure            160481..160537
                        /notes="Unresolved simple sequence repeat."

ORIGIN
Query Match          86.1%; Score 19.8; DB 9; Length 164494;
Best Local Similarity 91.3%; Pred.No. 81;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  TGAGAGGCGCAGTAGAAGCTTAG 23
        ||||| ||| ||||| |||||
Db       27851 TGAGAGACAGCAGAGGCTTAG 27829

RESULT 45
CT010447
LOCUS      CT010447          270432 bp    DNA    linear    HTG 11-AUG-2005
DEFINITION Mus musculus chromosome 13 clone RP23-15K11, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
CT010447
ACCESSION CT010447.2    GI:72534182
VERSION   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 270432)  
 McLaren, S.  
 Direct Submission  
 Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Aug 12, 2005 this sequence version replaced gi:71142025.

## ----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)

----- Project Information

Center project name: bM15K11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 267600 bases at least Q40

Consensus quality: 268187 bases at least Q30

Consensus quality: 268483 bases at least Q20

Insert size: 265332; sum-of-contigs

Insert size: 250336; 4.8% error; agarose-fp

Quality coverage: 7.49x in Q20 bases; sum-of-contigs Quality

coverage: 8.20x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 42623: contig of 42623 bp in length
* 42624 42723: gap of 100 bp
* 42724 76065: contig of 33342 bp in length
* 76066 76165: gap of 100 bp
* 76166 107675: contig of 31510 bp in length
* 107676 107776: gap of 101 bp
* 107777 127057: contig of 19281 bp in length
* 127058 127157: gap of 100 bp
* 127158 179315: contig of 52158 bp in length
* 179316 179415: gap of 100 bp
* 179416 190326: contig of 10911 bp in length
* 190327 190426: gap of 100 bp
* 190427 193918: contig of 3492 bp in length
* 193919 194019: gap of 101 bp
* 194020 203029: contig of 9010 bp in length
* 203030 203129: gap of 100 bp
* 203130 227389: contig of 24260 bp in length
* 227390 227489: gap of 100 bp
* 227490 270432: contig of 42943 bp in length.
```

## FEATURES

## Location/Qualifiers

## Source

```

1..270432
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="13"
/clone_lib="RPCI-23"
/clone="RP23-15K11"
1..42623
/note="assembly_fragment:01509
fragment_chain:1"
42724..76065
/note="assembly_fragment:00687
fragment_chain:1"
76166..107676
/note="assembly_fragment:00371
fragment_chain:1"
107777..127057
/note="assembly_fragment:00182

```

```

misc_feature      fragment chain:1"
127158..179315
/notes="assembly_fragment:02008
fragment_chain:1"
misc_feature      179416..190326
/notes="assembly_fragment:00055
fragment_chain:1"
misc_feature      190427..193918
/notes="assembly_fragment:00018"
194019..203029
/notes="assembly_fragment:02581"
misc_feature      203130..227389
/notes="assembly_fragment:02583"
misc_feature      227490..270432
/notes="assembly_fragment:01065
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      86.1%; Score 19.8; DB 14; Length 270432;
Best Local Similarity 91.3%; Pred.No. 78;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
||||| ||| ||||| |||||
Db 216240 TGAGAGACAGCAGAGAGCTTAG 216262

RESULT 46
CT010447/c
LOCUS             CT010447             270432 bp    DNA     linear     HTG 11-AUG-2005
DEFINITION       Mus musculus chromosome 13 clone RP23-15K11, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION        CT010447
VERSION          CT010447.2 GI:72534182
KEYWORDS         HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 270432)
McLaren.S.
Direct Submission
Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 12, 2005 this sequence version replaced gi:71142025.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
----- Project Information
Center project name: bM15K11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 267600 bases at least Q40
Consensus quality: 268187 bases at least Q30
Consensus quality: 268483 bases at least Q20
Insert size: 269532; sum-of-contigs
Insert size: 250336; 4.8% error; agarose-fp
Quality coverage: 7.49x in Q20 bases; sum-of-contigs Quality
coverage: 8.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 42623: contig of 42623 bp in length
* 42624 42723: gap of 100 bp
* 42724 76065: contig of 33342 bp in length
* 76066 76165: gap of 100 bp
* 76166 107675: contig of 31510 bp in length
* 107676 107776: gap of 101 bp
* 107777 127057: contig of 19281 bp in length
* 127058 127157: gap of 100 bp
* 127158 179315: contig of 52158 bp in length
* 179316 179415: gap of 100 bp
* 179416 190326: contig of 10911 bp in length
* 190327 190426: gap of 100 bp
* 190427 193918: contig of 3492 bp in length
* 193919 194019: gap of 101 bp
* 194020 203029: contig of 9010 bp in length
* 203030 203129: gap of 100 bp
* 203130 227389: contig of 24260 bp in length
* 227390 227489: gap of 100 bp
* 227490 270432: contig of 42943 bp in length.

FEATURES
            Location/Qualifiers
            1..270432
             /organism="Mus musculus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10090"
             /chromosome="13"
             /clone_lib="RP23-15K11"
             /clone_lib="RPC1-23"
             1..42623
              /notes="assembly_fragment:01509
              fragment_chain:1"
              42724..76065
               /notes="assembly_fragment:00687
               fragment_chain:1"
               76166..107676
                /notes="assembly_fragment:00371
                fragment_chain:1"
                107777..127057
                 /notes="assembly_fragment:00182
                 fragment_chain:1"
                 127158..179315
                  /notes="assembly_fragment:02008
                  fragment_chain:1"
                  179416..190326
                   /notes="assembly_fragment:00055
                   fragment_chain:1"
                   190427..193918
                    /notes="assembly_fragment:00018"
                    194019..203029
                     /notes="assembly_fragment:02581"
                     203130..227389
                      /notes="assembly_fragment:02583"
                      227490..270432
                       /notes="assembly_fragment:01065
                       clone_end:SP6
                       vector_side:right"

ORIGIN
Query Match      86.1%; Score 19.8; DB 14; Length 270432;
Best Local Similarity 91.3%; Pred.No. 78;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
||||| ||| ||||| |||||
Db 238550 TGAGAGACAGCAGAGAGCTTAG 238528

RESULT 47
AC110439/c
LOCUS             AC110439             319008 bp    DNA     linear     HTG 11-OCT-2002
DEFINITION       Rattus norvegicus clone CH230-287L21, *** SEQUENCING IN PROGRESS
***.
ACCESSION        AC110439

```

```

AC110439.4 GI:23604158
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 319008)
Munry,D,Marie., Metzker,M, Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Greggeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,W.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plapper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puaoro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,B., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Welas,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 319008)
Worley,K.C.
Direct Submission
Submitted (13-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 319008)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21738108.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQHE
Center clone name: CH230-287L21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 169602 bases at least Q40
Consensus quality: 172273 bases at least Q30
Consensus quality: 173833 bases at least Q20
Estimated insert size: 175232; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 319008: contig of 319008 bp in length.
-----
FEATURES
source
1..319008
Location/Qualifiers
1..319008
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-287L21"
723..1636
/note="clone boundary"
clone_end:Sp6
site:MboI
end sequence:RXAHS71TV"
complement(183223..184055)
/note="clone boundary"
clone_end:T7
site:MboI
end sequence:RXAHS71TV"
184135..185293
/note="wgs_end_extension"
clone_end:T7"
31971..319008
/note="wgs_end_extension"
clone_end:T7"
-----
ORIGIN
Query Match 86.1%; Score 19.8; DB 14; Length 319008;
Best Local Similarity 91.3%; Pred. No. 77;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAGAGGCGACTAGAAAGCTTAG 23
|||||
Db 95734 TGAGAGGCGAGCAGACGCTTAG 95712
|||||
RESULT 48

```

```

AC009610      AC009610      60036 bp      DNA      linear      HTG 24-DEC-2000
LOCUS          Homo sapiens chromosome 4 clone RP11-396 map 4, LOW-PASS SEQUENCE
DEFINITION     SAMPLING.
ACCESSION      AC009610
VERSION        AC009610.2 GI:11991318
KEYWORDS       HTG; HTGS_PHASE0.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 60036)
AUTHORS        Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE          Homo sapiens chromosome 4, clone RP11-396
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 60036)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
               Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
               Caule,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
               Cooke,P., Dearellano,K., Depayre,S., Devon,K., Dewar,K.,
               Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
               Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
               Karas,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
               Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
               Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
               Meidrum,J., Mollia,M., Morris,W., Morrow,J., Mychaleckyj,J.,
               Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
               Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
               Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
               Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
               Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On Dec 24, 2000 this sequence version replaced gi:5801798.
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Project name: I2010
               Center project name: 39_D_6
               -----
               * NOTE: This record contains 73 individual
               * sequencing reads that have not been assembled into
               * contigs. Runs of N are used to separate the reads
               * and the order in which they appear is completely
               * arbitrary. Low-pass sequence sampling is useful for
               * identifying clones that may be gene-rich and allows
               * overlap relationships among clones to be deduced.
               * However, it should not be assumed that this clone
               * will be sequenced to completion. In the event that
               * the record is updated, the accession number will
               * be preserved.
               *
               * 1 751: contig of 751 bp in length
               * 752 851: gap of 100 bp
               * 852 1590: contig of 739 bp in length
               * 1591 1690: gap of 100 bp
               * 1691 2381: contig of 691 bp in length
               * 2382 2481: gap of 100 bp
               * 2482 3211: contig of 730 bp in length
               * 3212 3311: gap of 100 bp
               * 3312 4041: contig of 730 bp in length
               * 4042 4141: gap of 100 bp
               * 4142 4890: contig of 749 bp in length
               * 4891 4991: gap of 100 bp
               * 4991 5702: contig of 712 bp in length
               * 5703 5802: gap of 100 bp
               *
               * 5803 contig of 701 bp in length
               * 5804 gap of 100 bp
               * 5805 contig of 693 bp in length
               * 5806 gap of 100 bp
               * 5807 contig of 728 bp in length
               * 5808 gap of 100 bp
               * 5809 contig of 711 bp in length
               * 5810 gap of 100 bp
               * 5811 contig of 730 bp in length
               * 5812 gap of 100 bp
               * 5813 contig of 741 bp in length
               * 5814 gap of 100 bp
               * 5815 contig of 720 bp in length
               * 5816 gap of 100 bp
               * 5817 contig of 730 bp in length
               * 5818 gap of 100 bp
               * 5819 contig of 752 bp in length
               * 5820 gap of 100 bp
               * 5821 contig of 713 bp in length
               * 5822 gap of 100 bp
               * 5823 contig of 717 bp in length
               * 5824 gap of 100 bp
               * 5825 contig of 699 bp in length
               * 5826 gap of 100 bp
               * 5827 contig of 737 bp in length
               * 5828 gap of 100 bp
               * 5829 contig of 737 bp in length
               * 5830 gap of 100 bp
               * 5831 contig of 732 bp in length
               * 5832 gap of 100 bp
               * 5833 contig of 733 bp in length
               * 5834 gap of 100 bp
               * 5835 contig of 729 bp in length
               * 5836 gap of 100 bp
               * 5837 contig of 752 bp in length
               * 5838 gap of 100 bp
               * 5839 contig of 713 bp in length
               * 5840 gap of 100 bp
               * 5841 contig of 717 bp in length
               * 5842 gap of 100 bp
               * 5843 contig of 699 bp in length
               * 5844 gap of 100 bp
               * 5845 contig of 737 bp in length
               * 5846 gap of 100 bp
               * 5847 contig of 732 bp in length
               * 5848 gap of 100 bp
               * 5849 contig of 710 bp in length
               * 5850 gap of 100 bp
               * 5851 contig of 733 bp in length
               * 5852 gap of 100 bp
               * 5853 contig of 729 bp in length
               * 5854 gap of 100 bp
               * 5855 contig of 752 bp in length
               * 5856 gap of 100 bp
               * 5857 contig of 752 bp in length
               * 5858 gap of 100 bp
               * 5859 contig of 681 bp in length
               * 5860 gap of 100 bp
               * 5861 contig of 726 bp in length
               * 5862 gap of 100 bp
               * 5863 contig of 726 bp in length
               * 5864 gap of 100 bp
               * 5865 contig of 692 bp in length
               * 5866 gap of 100 bp
               * 5867 contig of 735 bp in length
               * 5868 gap of 100 bp
               * 5869 contig of 714 bp in length
               * 5870 gap of 100 bp
               * 5871 contig of 741 bp in length
               * 5872 gap of 100 bp
               * 5873 contig of 722 bp in length
               * 5874 gap of 100 bp
               * 5875 contig of 713 bp in length
               * 5876 gap of 100 bp
               * 5877 contig of 731 bp in length
               * 5878 gap of 100 bp
               * 5879 contig of 725 bp in length
               * 5880 gap of 100 bp
               * 5881 contig of 766 bp in length
               * 5882 gap of 100 bp
               * 5883 contig of 712 bp in length
               * 5884 gap of 100 bp
               * 5885 contig of 724 bp in length
               * 5886 gap of 100 bp
               * 5887 contig of 732 bp in length
               * 5888 gap of 100 bp
               * 5889 contig of 696 bp in length
               * 5890 gap of 100 bp
               * 5891 contig of 749 bp in length
               * 5892 gap of 100 bp
               * 5893 contig of 693 bp in length
               * 5894 gap of 100 bp

```

* 36172	36271: gap of 100 bp				
* 36272	36997: contig of 726 bp in length				
* 36998	37097: gap of 100 bp				
* 37098	37829: contig of 732 bp in length				
* 37830	37929: gap of 100 bp				
* 37930	38677: contig of 748 bp in length				
* 38678	38777: gap of 100 bp				
* 38778	39508: contig of 731 bp in length				
* 39509	39508: contig of 100 bp				
* 39609	40340: contig of 732 bp in length				
* 40341	40440: gap of 100 bp				
* 40441	41177: contig of 737 bp in length				
* 41178	41277: gap of 100 bp				
* 41278	41969: contig of 692 bp in length				
* 41970	42069: gap of 100 bp				
* 42070	42809: contig of 740 bp in length				
* 42810	42909: gap of 100 bp				
* 42910	43613: contig of 704 bp in length				
* 43614	43713: gap of 100 bp				
* 43714	44442: contig of 729 bp in length				
* 44443	44542: gap of 100 bp				
* 44543	45258: contig of 716 bp in length				
* 45259	45358: gap of 100 bp				
* 45359	46082: contig of 724 bp in length				
* 46083	46182: gap of 100 bp				
* 46183	46922: contig of 740 bp in length				
* 46923	47022: gap of 100 bp				
* 47023	47596: contig of 674 bp in length				
* 47697	47796: gap of 100 bp				
* 47797	48549: contig of 753 bp in length				
* 48550	48649: gap of 100 bp				
* 48650	49370: contig of 721 bp in length				
* 49371	49470: gap of 100 bp				
* 49471	50201: contig of 731 bp in length				
* 50202	50301: gap of 100 bp				
* 50302	51047: contig of 746 bp in length				
* 51048	51147: gap of 100 bp				
* 51148	51894: contig of 737 bp in length				
* 51885	51984: gap of 100 bp				
* 51985	52692: contig of 708 bp in length				
* 52693	52792: gap of 100 bp				
* 52793	53511: contig of 719 bp in length				
* 53512	53611: gap of 100 bp				
* 53612	54243: contig of 632 bp in length				
* 54244	54343: gap of 100 bp				
* 54344	55106: contig of 763 bp in length				
* 55107	55206: gap of 100 bp				
* 55207	55937: contig of 731 bp in length				
* 55938	56037: gap of 100 bp				
* 56038	56738: contig of 701 bp in length				
* 56739	56838: gap of 100 bp				
* 56839	57546: contig of 708 bp in length				
* 57547	57646: gap of 100 bp				
* 57647	58332: contig of 686 bp in length				
* 58333	58432: gap of 100 bp				

Query Match

Best Local Similarity

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

2

GAGAGGCGAGTAAAGCTTAG 23

Db

389

GAGAGGAGTAGAAGCCTAG 410

RESULT 49

AC138779/c

LOCUS

AC138779 Homo sapiens BAC clone RP11-60115 from 4, complete sequence.

DEFINITION

AC138779

ACCESSION

AC138779.2 GI:27802069

VERSION

HTG.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo  
1 (bases 1 to 132003)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
9847074  
2 (bases 1 to 132003)  
Van Brunt, A. and Bielicki, L.  
The sequence of Homo sapiens BAC clone RP11-60115  
Unpublished (2001)  
3 (bases 1 to 132003)  
Waterston, R.H.  
Direct Submission  
Submitted (18-JAN-2003) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 132003)  
Waterston, R.H.  
Direct Submission  
Submitted (17-JAN-2003) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 132003)  
Waterston, R.H.  
Direct Submission  
Submitted (18-JAN-2003) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 132003)  
Waterston, R.H.  
Direct Submission  
Submitted (27-JAN-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 18, 2003 this sequence version replaced gi:27764832.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
----- Center project name: H\_NH060115  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is not the entire insert of the clone. This clone is overlapped by AC138607 and AC098587.

## FEATURES

Location/Qualifiers  
1..132003  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-601115"  
/clone\_lib="RPC1-11"  
56..95  
/rpt\_family="TATG)n"  
240..291  
/rpt\_family="T-rich"  
292..839  
/rpt\_family="L1"  
1460..11776  
/rpt\_family="MER2\_type"  
1857..1895  
/rpt\_family="MaLR"  
1968..2091  
/rpt\_family="Alu"  
2093..2219  
/rpt\_family="GA-rich"  
2692..3045  
/rpt\_family="MaLR"  
3122..3214  
/rpt\_family="L1"  
3215..3525  
/rpt\_family="Alu"  
3526..4009  
/rpt\_family="L1"  
4010..4178  
/rpt\_family="Alu"  
4193..4795  
/rpt\_family="L1"  
4796..4815  
/rpt\_family="(TTAAA)n"  
4816..5821  
/rpt\_family="L1"  
5833..6023  
/rpt\_family="L1"  
6024..6321  
/rpt\_family="Alu"  
6322..6731  
/rpt\_family="L1"  
7006..7138  
/rpt\_family="MER1\_type"  
7284..7810  
/rpt\_family="ERV1"  
9847..10124  
/rpt\_family="Alu"  
10878..11173  
/rpt\_family="Alu"  
11577..11883  
/rpt\_family="Alu"  
11934..12684  
/rpt\_family="L1"  
12688..12978  
/rpt\_family="Alu"  
13285..13584  
/rpt\_family="Alu"  
13589..13698  
/rpt\_family="MIR"  
13835..13921  
/rpt\_family="CR1"  
14786..15097  
/rpt\_family="Alu"  
15295..15589

repeat\_region  
16100..16247  
/rpt\_family="U3"  
repeat\_region  
16267..16456  
/rpt\_family="Alu"  
repeat\_region  
16457..16520  
/rpt\_family="ERV1"  
repeat\_region  
16542..16639  
/rpt\_family="Alu"  
repeat\_region  
16640..16712  
/rpt\_family="U3"  
repeat\_region  
16727..17035  
/rpt\_family="Alu"  
repeat\_region  
17219..18629  
/rpt\_family="ERV1"  
repeat\_region  
19055..19296  
/rpt\_family="Alu"  
repeat\_region  
19423..19620  
/rpt\_family="L1"  
repeat\_region  
19699..19857  
/rpt\_family="MIR"  
repeat\_region  
20605..20883  
/rpt\_family="Alu"  
repeat\_region  
21043..21487  
/rpt\_family="L2"  
repeat\_region  
21778..21934  
/rpt\_family="MaLR"  
repeat\_region  
22047..22206  
/rpt\_family="Alu"  
repeat\_region  
22317..22410  
/rpt\_family="L2"  
repeat\_region  
22486..22929  
/rpt\_family="ERV1"  
repeat\_region  
24659..25077  
/rpt\_family="MaLR"  
repeat\_region  
25606..25674  
/rpt\_family="MaLR"  
repeat\_region  
25883..25960  
/rpt\_family="Alu"  
repeat\_region  
25964..26009  
/rpt\_family="polypurine"  
repeat\_region  
26010..26099  
/rpt\_family="GA-rich"  
repeat\_region  
26100..26396  
/rpt\_family="Alu"  
repeat\_region  
26397..26472  
/rpt\_family="GA-rich"  
repeat\_region  
26578..26786

Query Match 81.7%; Score 18.8; DB 8; Length 132003;

Best Local Similarity 90.9%; Pred. No. 2.5e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAAGCAGTAGAAGCTTAG 23

Db 92407 GAGAAAGCAGTAGAAGCTTAG 92386

RESULT 50

AC026580

LOCUS

DEFINITION

AC026580

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

164084 bp DNA linear HTG 20-OCT-2001  
Homo sapiens chromosome 4 clone RP11-39D6 map 4, WORKING DRAFT  
SEQUENCE, 17 unordered pieces.  
AC026580.2 GI:16303507  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 164084)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.







Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhsawa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, G., Newton, N., Nguyen, N., Norris, S., Nwackelmele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Paaternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plöpper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Unpublished  
2 (bases 1 to 191119)  
Worley, K.C.  
Direct Submission  
Submitted (21-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 191119)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23907920.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GWY  
Center clone name: CH230-30903  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 176412 bases at least Q40

Consensus quality: 179015 bases at least Q30  
Consensus quality: 180610 bases at least Q20  
Estimated insert size: 181867; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 189365: contig of 189365 bp in length  
\* 189366 189465: gap of unknown length  
\* 189466 191119: contig of 1654 bp in length.

#### FEATURES

Location/Qualifiers

source

1..191119  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-30903"  
150921..179194  
/note="clone boundary"  
clone\_end:Sp5  
site:  
end\_sequence:BZ124947"  
159049..160115  
/note="wgs end\_extension"  
clone\_end:Sp6"  
189365..189465  
/estimated\_length=unknown

#### misc\_feature

#### misc\_feature

#### gap

#### ORIGIN

Query Match 81.7%; Score 18.8; DB 14; Length 191119;  
Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GAGAAAGGCGAGTAGAAAGCTTAG 23  
|||||  
Db 61950 GAGAAAGGCGAGTAGAAAGCATGG 61929

#### RESULT 53

#### AC117689

#### LOCUS

#### DEFINITION

#### AC117689

#### ACCESSION

#### VERSION

#### KEYWORDS

#### ORGANISM

#### SOURCE

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS</

McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 200562)

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**

Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Faro, S., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasany, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (09-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 200562)

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**

Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Faro, S., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasany, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (17-JAN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jan 17, 2004 this sequence version replaced gi:39573811.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
----- Center project name: L24076

Center clone name: 446\_I\_19  
-----  
**FEATURES**  
**source**  
1. 200562  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/map="12"  
/clone="RP23-446I19"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. 569  
note="wgs\_end\_extension"  
clone\_end:SP6"  
570. 575  
/notes="clone boundary"  
clone\_end:SP6  
site:ECORI"  
615. 972  
/rpt\_family="MER76"  
complement(1851..1927)  
/rpt\_family="ID\_B1"  
2002. 2040  
/rpt\_family="AT\_rich"  
2904. 3075  
/rpt\_family="L1MA7"  
complement(3406..9520)  
/rpt\_family="L1F"  
9542. 9579  
/rpt\_family="AT\_rich"  
9858. 9896  
/rpt\_family="(TAAAA)n"  
9989. 10209  
/rpt\_family="B4A"  
complement(10668..11339)  
/rpt\_family="Lx2"  
11365. 11527  
/rpt\_family="Lx2"  
complement(11531..11709)  
/rpt\_family="L1\_MM"  
11710. 12632  
/rpt\_family="Lx2"  
12671. 12699  
/rpt\_family="(CAA)n"  
complement(13503..13652)  
/rpt\_family="MER58"  
14354. 15308  
/rpt\_family="RLTR13C"  
15335. 15366  
/rpt\_family="(TTTTG)n"  
15991. 16308  
/rpt\_family="Lx9"  
16963. 16985  
/rpt\_family="(CAAA)n"  
17157. 18470  
/rpt\_family="Lx9"  
18856. 18987  
/rpt\_family="(GGAA)n"  
19564. 19611  
/rpt\_family="(TA)n"  
20821. 20864  
/rpt\_family="(CAAAA)n"  
complement(20936..21231)  
/rpt\_family="L1\_MM"  
21233. 21388  
/rpt\_family="L1\_MM"  
21844. 21889  
/rpt\_family="(TTTTG)n"  
complement(22130..22306)  
/rpt\_family="ID\_B1"  
complement(23225..23466)  
/rpt\_family="URRIA"  
complement(24053..24218)  
repeat\_region

```

/rpt family="Lx2"
2421g .24988
/rpt family="Lx"
25227 .31255
/rpt family="L1_MM"
31398 .31426
/rpt family="(TC)n"
3142g .31468
/rpt family="(TA)n"
31580 .31651
/rpt family="(T)n"
complement(31661 .31798)
/rpt family="B1_MM"
complement(31877 .32017)
/rpt family="URR1B"
32031 .32052
/rpt family="AT_rich"
32201 .32300
/rpt family="(T)n"
33621 .33681
/rpt family="(CA)n"
complement(33745 .33880)
/rpt family="B4g"
33954 .33982
/rpt family="AT_rich"
34165 .35945
/rpt family="Lx2"
36660 .36695

Query Match      81.7%; Score 18.8; DB 9; Length 200562;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGAGTAGAAGCTTA 22
||||| ||||| ||||| |||||
Db 140449 TGAGAGGCGAGTAGAAGCTTA 140470

RESULT 54
AC107536/c
LOCUS
DEFINITION
AC107536 239438 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-211116, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC107536
AC107536.6 GI:30581446
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
REFERENCE
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Margum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.-L.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Tabor, P., Taylor, C.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 239438)
Worley, K.C.
Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239438)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24942470.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKPU
Center clone name: CH230-211116
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 21615 bases at least Q40
Consensus quality: 21613 bases at least Q30
Consensus quality: 21877 bases at least Q20
Estimated insert size: 226897; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 5696: contig of 5696 bp in length  
 \* 5697 5796: gap of unknown length  
 \* 5797 237972: contig of 232176 bp in length  
 \* 237973 238072: gap of unknown length  
 \* 238073 239438: contig of 1366 bp in length.

## FEATURES

source

1..239438  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-211L16"

misc\_feature

1..1924

/note="wgs\_end\_extension"

clone\_end:Sp6"

1975..5696

misc\_feature

/note="wgs\_end\_extension"

clone\_end:Sp6"

complement(2994..3834)

/note="clone\_boundary"

clone\_end:Sp6"

site:EcoRI

end\_sequence:RWBLD68TVB"

5697..5796

/estimated\_length=unknown

misc\_feature

5797..7308

/note="wgs\_contig"

169595..172710

/note="wgs\_contig"

189620..191165

/note="wgs\_contig"

193829..195622

/note="wgs\_contig"

complement(230967..231795)

/note="clone\_boundary"

clone\_end:T7"

site:EcoRI

end\_sequence:RWBLD68TVB"

237973..238072

/estimated\_length=unknown

## ORIGIN

Query Match 81.7%; Score 18.8; DB 14; Length 239438;  
 Best Local Similarity 90.9%; Pred. NO. 2.4e-02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGCTTAG 23

|||||

Db 90230 GAGAGGCGAGTAGAAGCTTTG 90209

RESULT 55

AC128575/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-149J17, WORKING DRAFT SEQUENCE, 3

AC128575

unordered pieces.

AC128575.2 GI:23907831

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 239808)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Auguiano,D., Ayalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biewalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaý,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J., Maheehwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Napua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Popper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojase,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajds,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Vallar,R., Vera,V., Villaeana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 239808)

REFERENCE

Worley,K.C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 239808)

REFERENCE

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 12, 2002 this sequence version replaced gi:21909348.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GROQ  
Center clone name: CH230-149J17  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 166535 bases at least Q40  
Consensus quality: 168913 bases at least Q30  
Consensus quality: 170484 bases at least Q20  
Estimated insert size: 168276; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 237163: contig of 237163 bp in length  
\* 237164 237263: gap of unknown length  
\* 237264 238491: contig of 1228 bp in length  
\* 238492 238591: gap of unknown length  
\* 238592 239808: contig of 1217 bp in length.

## FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

gap

gap

## ORIGIN

Query Match 81.7%; Score 18.8; DB 14; Length 239808;  
Best Local Similarity 90.9%; Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GAGAGGCGAGTGAAGCTTAG 23  
|||||  
DB 184147 GAGAGGCGAGTGAAGCATGG 184126  
|||||

## RESULT 56

AC128356

LOCUS

DEFINITION

AC128356

AC128356.3

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC128356 243280 bp DNA linear HTG 15-NOV-2002  
Rattus norvegicus clone CH230-146J15, WORKING DRAFT SEQUENCE, 4  
unordered pieces.  
AC128356 GI:25007899  
HTG: HTGS PHASE1: HTGS DRAFT; HTGS FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.  
1 (bases 1 to 243280)  
Muzny D, Maris, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,  
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, S., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,  
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, L., Loulseg, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokelemeh, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleciyk, R., Woodden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 243280)

Worley, K. C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 243280)

REFERENCE





```
Query Match      80.0%; Score 18.4; DB 8; Length 3310;
Best Local Similarity 95.0%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTAGAAGCT 20
          | | | | | | | | | |
Db      2348 TAAGAAGGCAGTAGAAGCT 2329

RESULT 58
AC009010
LOCUS      Homo sapiens chromosome 5 clone XXp1-1352A1, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
ACCESSION AC009010
VERSION    AC009010.5 GI:11178047
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 60634)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 5
            Unpublished
            2 (bases 1 to 60634)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Nov 15, 2000 this sequence version replaced gi:9964734.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1189015, H231
Center clone name: XXp1-1352A1
-----
Summary Statistics
Consensus quality: 50043 bases at least Q40
Consensus quality: 55954 bases at least Q30
Consensus quality: 57601 bases at least Q20
Estimated insert size: 8000; pulse field gel estimation
Quality coverage: 2.82 in Q20 bases; pulse field gel estimation
Quality coverage: 3.77 in Q20 bases; sun-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1297: contig of 1297 bp in length
* 1298 1397: gap of unknown length
* 1398 2981: contig of 1584 bp in length
* 2982 3081: gap of unknown length
* 3082 5483: contig of 2402 bp in length
* 5484 5583: gap of unknown length
* 5584 11308: contig of 5725 bp in length
* 11309 11408: gap of unknown length
* 11409 16459: contig of 5051 bp in length
* 16460 16559: gap of unknown length
* 16560 25749: contig of 9190 bp in length
* 25750 25849: gap of unknown length
* 25850 34859: contig of 9010 bp in length
* 34860 34959: gap of unknown length
* 34960 60634: contig of 25675 bp in length.
            Location/Qualifiers

FEATURES
            source
            1..60791
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosomes="5"

source
            1..60634
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="XXp1-1352A1"
            1298..1397
            /estimated_length=unknown
            2982..3081
            /estimated_length=unknown
            5484..5583
            /estimated_length=unknown
            11309..11408
            /estimated_length=unknown
            16460..16559
            /estimated_length=unknown
            25750..25849
            /estimated_length=unknown
            34860..34959
            /estimated_length=unknown

ORIGIN
Query Match      80.0%; Score 18.4; DB 14; Length 60634;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTAGAAGCT 20
          | | | | | | | | | |
Db      35406 TAAGAAGGCAGTAGAAGCT 35425

RESULT 59
AC004749/c
LOCUS      Homo sapiens chromosome 5, P1 clone 632A8 (LBNL H29), complete
DEFINITION sequence.
ACCESSION AC004749 AC000974 AC000975 AC000977 AC001489 AC002208
            L48484 L49048 L77054 L77055 L77056 L77057 L78758 L78759
            L81379 L81380 L81628 L81795 L81796 L81797 L81798
            AC004749.1 GI:3157918
VERSION    HTG.
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE  1 (bases 1 to 60791)
            Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
            Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
            Rojeski, H., Subramanian, S. and Martin, C.H.
            Sequencing of human chromosome 5
            Unpublished
            2 (bases 1 to 60791)
            Ricke, D.O.
            Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
            Unpublished
            3 (bases 1 to 60791)
            Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
            Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
            Rojeski, H., Subramanian, S. and Martin, C.H.
            Direct Submission
            Submitted (27-MAY-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
            Sequence submitted by:
            DOE Joint Genome Institute.
            Location/Qualifiers

FEATURES
            source
            1..60791
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosomes="5"
```

```
/map="5c"
/clone="f32a8"
/note="DBNL H29"
1. .60791
/rpt_note="(T) 27"
/rpt_type=tandem
/rpt_unit="t"
453. .757
/rpt_family="MER1B"
complement(1157. .1283)
/note="GRAIL 2 excellent exon, frame 0"
complement(1177. .1282)
/note="93% identity mouse mp17d11.r1"
/db_xref="dbEST:AA288684"
complement(1882. .2044)
/rpt_family="Alu"
1891. .1913
/note="(T) 23"
/rpt_type=tandem
/rpt_unit="t"
5204. .5503
/rpt_family="Alu"
6376. .6912
/rpt_family="Alu"
8028. .8115
/note="GRAIL 2 excellent exon, frame 2"
complement(9142. .9288)
/rpt_family="MER5"
9356. .9674
/rpt_family="Alu"
9611. .10326
/rpt_family="L1"
complement(10569. .10869)
/rpt_family="Alu"
10859. .10943
/note="GRAIL 2 excellent exon, frame 1"
complement(11133. .11417)
/note="99% identity EST60905"
/db_xref="dbEST:AA352848"
complement(12263. .12538)
/rpt_family="Alu"
12276. .12296
/note="(T) 21"
/rpt_type=tandem
/rpt_unit="t"
complement(14099. .14403)
/rpt_family="Alu"
14892. .15199
/note="97% identity EST02082"
/db_xref="dbEST:M85566"
16203. .16362
/rpt_family="Alu"
17932. .18067
/rpt_family="MER42"
complement(18549. .18857)
/rpt_family="Alu"
18552. .18868
/note="(T) 21"
/rpt_type=tandem
/rpt_unit="t"
complement(19792. .20110)
/rpt_family="Alu"
complement(20826. .21152)
/rpt_family="Alu"
21162. .21549
/note="99% identity yu10e03.r1"
/db_xref="dbEST:H79907"
complement(21500. .21824)
/rpt_family="Alu"
22829. .23118
/rpt_family="Alu"
23149. .23457
/rpt_family="Alu"

repeat_region complement(23659. .23973)
/rpt_family="Alu"
24620. .25456
/rpt_family="LTR1"
25651. .26710
/rpt_family="Alu"
25933. .26038
/rpt_family="Alu"
26390. .28695
/rpt_family="Alu"
complement(26796. .26904)
/rpt_family="MIR"
complement(27505. .27616)
/note="99% identity"
/db_xref="dbEST:D63193"
complement(29907. .30188)
/rpt_family="Alu"
complement(31287. .31532)
/rpt_family="Alu"
31649. .31828
/rpt_family="MER46"
31831. .31976
/rpt_family="Alu"
complement(32558. .32880)
/rpt_family="Alu"
complement(32887. .33189)
/rpt_family="Alu"
33219. .33558
/rpt_family="Alu"
33559. .33843
/rpt_family="Alu"
complement(34684. .34783)
/rpt_family="L1"
35259. .35558
/rpt_family="Alu"
35561. .35867
/rpt_family="Alu"
36176. .36203
/note="(AC) 14"
/rpt_type=tandem
/rpt_unit="ac"
complement(37312. .42768)
/rpt_family="L1"
complement(43077. .43389)
/rpt_family="Alu"
43834. .44199
/rpt_family="Alu"
44128. .44446
/rpt_family="Alu"
44424. .44448
/note="(A) 20"
/rpt_type=tandem
/rpt_unit="a"
complement(44448. .44837)
/rpt_family="Alu"
complement(44895. .45025)
/rpt_family="Alu"
45828. .46140
/rpt_family="Alu"
complement(46222. .46303)
/rpt_family="MER42C"
46466. .46636
/rpt_family="MER44C"
complement(46987. .47304)
/rpt_family="Alu"
complement(47536. .47862)
/rpt_family="Alu"
48111. .48231
/rpt_family="MIR"
48705. .54852
/rpt_family="L1"
complement(55742. .56085)
/rpt_family="Alu"
```

Search completed: January 27, 2006, 23:13:39  
Job time : 915.685 secs

```
repeat_region 55940..55959
/note="(T)20"
/rpt_type=tandem
/rpt_unit="t"
repeat_region complement(56109..56261)
/rpt_family="MER20"
misc_feature complement(56361..56800)
/note="97% identity yp96g11.r1"
/db_xref="dbEST:R92125"
repeat_region 56771..56875
/rpt_family="MLT1F"
repeat_region complement(56884..57190)
/rpt_family="Alu"
repeat_region 58783..58908

Query Match 80.0%; Score 18.4; DB 8; Length 60791;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGCAGTAGAAGCT 20
DB 20410 TAAGAGGCGCAGTAGAAGCT 20391

RESULT 60
AC009016 79041 bp DNA linear PRI 08-NOV-2000
LOCUS Homo sapiens chromosome 5 clone P1_889E7, complete sequence.
DEFINITION AC009016
ACCESSION AC009016
VERSION AC009016.3 GI:11120759
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 79041)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 79041)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 79041)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 8, 2000 this sequence version replaced gi:6862871.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
source
1..79041
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="P1_889E7"

ORIGIN
Query Match 80.0%; Score 18.4; DB 8; Length 79041;
Best Local Similarity 95.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGCAGTAGAAGCT 20
DB 65068 TAAGAGGCGCAGTAGAAGCT 65087
```

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 824,191 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattgatgggttatcttcc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_ov.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_sts.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	CS109171 Sequence
2	21	100.0	21	6	AR353653 Sequence
3	21	100.0	21	6	AX817898 Sequence
4	21	100.0	1323	6	AR353662 Sequence
5	21	100.0	1323	6	AX817899 Sequence
6	21	100.0	1378	6	AR353664 Sequence
7	21	100.0	1378	6	AX817901 Sequence
8	21	100.0	1379	6	AR353657 Sequence
9	21	100.0	1379	6	AR353663 Sequence
10	21	100.0	1379	6	AR353666 Sequence
11	21	100.0	1379	6	AX817894 Sequence
12	21	100.0	1379	6	AX817900 Sequence
13	21	100.0	1379	6	AR353658 Sequence
14	21	100.0	1384	6	AR353658 Sequence
15	21	100.0	1384	6	AX817895 Sequence
16	21	100.0	1384	6	AX817895 Sequence
17	21	100.0	1384	6	AX817897 Sequence
18	21	100.0	1385	6	AR353659 Sequence

C 19	21	100.0	1385	6	AX817896	AX817896 Sequence
C 20	21	100.0	1390	6	AR353661	AR353661 Sequence
C 21	21	100.0	1390	6	AX817898	AX817898 Sequence
C 22	21	100.0	1393	6	AR353665	AR353665 Sequence
C 23	21	100.0	1393	6	AX817902	AX817902 Sequence
C 24	21	100.0	1447	6	AR353667	AR353667 Sequence
C 25	21	100.0	1447	6	AX817904	AX817904 Sequence
C 26	21	100.0	1731	6	CQ647508	CQ647508 Sequence
C 27	21	100.0	1731	6	CQ647510	CQ647510 Sequence
C 28	21	100.0	1734	6	AX608629	AX608629 Sequence
C 29	21	100.0	1803	6	AR353656	AR353656 Sequence
C 30	21	100.0	1803	6	AX817892	AX817892 Sequence
C 31	21	100.0	3797	1	AB027569	AB027569 Streptoc
C 32	21	100.0	10430	1	AE006575	AE006575 Streptoc
C 33	21	100.0	11992	1	AE010058	AE010058 Streptoc
C 34	21	100.0	21781	1	AE014229	AE014229 Streptoc
C 35	21	100.0	29072	6	AX602188	AX602188 Sequence
C 36	21	100.0	51454	1	AE014157	AE014157 Streptoc
C 37	21	100.0	110000	1	BA000034	Continuation (8 of
C 38	21	100.0	110000	1	BA000034	Continuation (9 of
C 39	21	100.0	110000	1	CP000003	Continuation (11 o
C 40	21	100.0	110000	1	CP000017	Continuation (11 o
C 41	21	100.0	110000	1	CP000056	Continuation (12 o
C 42	21	100.0	167050	1	SAG766847	AL766847 Streptoc
C 43	21	100.0	349980	6	CQ655069	CQ655069 Sequence
C 44	21	100.0	349980	6	AX954529	AX954529 Sequence
C 45	19.4	92.4	402	6	AX569143	AX569143 Sequence
C 46	19.4	92.4	1140	6	AR480287	AR480287 Sequence
C 47	19.4	92.4	1287	6	AR587350	AR587350 Sequence
C 48	19.4	92.4	2259	1	STREIPEPA	M81756 Streptococ
C 49	19.4	92.4	8195	6	CQ789003	CQ789003 Sequence
C 50	19.4	92.4	8195	6	AR218862	AR218862 Sequence
C 51	19.4	92.4	8195	6	BD003774	BD003774 Polynucle
C 52	19.4	92.4	10256	1	AE007418	AE007418 Streptoc
C 53	19.4	92.4	10320	1	AE008479	AE008479 Streptoc
C 54	19.4	92.4	20035	14	SPNEU1915	AL449937 Streptoc
C 55	19.4	92.4	349980	6	AX571763	AX571763 Sequence
C 56	18.4	87.6	158063	8	AP001046	AP001046 Homo sapi
C 57	18.4	87.6	178293	14	AC144901	AC144901 Sus acrofi
C 58	18.4	87.6	179556	14	AC018734	AC018734 Homo sapi
C 59	18.4	87.6	211897	14	AC118500	AC118500 Rattus no
C 60	18.4	87.6	340000	8	AP001751	AP001751 Homo sapi

## ALIGNMENTS

RESULT 1	CS109171	CS109171	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	CS109171	CS109171	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
DEFINITION	CS109171	CS109171	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
ACCESSION	CS109171	CS109171	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
VERSION	CS109171.1	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
KEYWORDS	CS109171.1	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
SOURCE	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
ORGANISM	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
REFERENCE	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
AUTHORS	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
TITLE	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
JOURNAL	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
FEATURES	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
source	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
synthetic construct	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
other sequences; artificial sequences.	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
1	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
Uhl, J.R., Cockerill, F.R., Alchinger, C. and Reiser, A.	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
Detection of group B streptococcus	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
Patent: Ep 1541697-A 2 15-JUN-2005;	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
Mayo Foundation for Medical Education and Research (US)	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
Location/Qualifiers	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
1..21	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
/organism="synthetic construct"	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
/mol_type="unassigned DNA"	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
/db_xref="taxon:32630"	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005
/note="Oligonucleotide"	CS109171	GI:68148020	Sequence 2 from Patent EP1541697.	21 bp	DNA	linear	PAT 22-JUN-2005

## ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTTATCTTCC 21  
Db 1 TGCATGATGGGTTATCTTCC 21

RESULT 2  
AR353653  
LOCUS AR353653 21 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 2 from patent US 6593093.  
ACCESSION AR353653  
VERSION AR353653.1 GI:33759686  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 2 15-JUL-2003;  
MAYO Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1..21  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred.No. 22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTTATCTTCC 21  
Db 1 TGCATGATGGGTTATCTTCC 21

RESULT 3  
AR353653  
LOCUS AR353653 21 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 2 from Patent EPI338656.  
ACCESSION AR353653  
VERSION AR353653.1 GI:39723076  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Uhl,J.R. and Cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 2 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
1..21  
Location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xrefs="taxon:32830"  
/note="Oligonucleotide"

ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred.No. 22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTTATCTTCC 21  
Db 1 TGCATGATGGGTTATCTTCC 21

RESULT 4  
AR353662/c  
LOCUS AR353662 1323 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 12 from patent US 6593093.  
ACCESSION AR353662

VERSION AR353662.1 GI:33759695  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1323)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 12 15-JUL-2003;  
MAYO Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1..1323  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1323;  
Best Local Similarity 100.0%; Pred.No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTTATCTTCC 21  
Db 193 TGCATGATGGGTTATCTTCC 173

RESULT 5  
AR353664/c  
LOCUS AR353664 1323 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 12 from Patent EPI338656.  
ACCESSION AR353664  
VERSION AR353664.1 GI:39723085  
KEYWORDS  
SOURCE Streptococcus sp. 'group A'  
ORGANISM Streptococcus sp. 'group A'  
REFERENCE 1  
AUTHORS Uhl,J.R. and Cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 12 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
1..1323  
Location/Qualifiers  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptai sequence from isolate no. 2"

ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1323;  
Best Local Similarity 100.0%; Pred.No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTTATCTTCC 21  
Db 193 TGCATGATGGGTTATCTTCC 173

RESULT 6  
AR353664/c  
LOCUS AR353664 1378 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 14 from patent US 6593093.  
ACCESSION AR353664  
VERSION AR353664.1 GI:33759697  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1378)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 14 15-JUL-2003;  
MAYO Foundation for Medical Education and Research; Rochester, MN

```
FEATURES
  source
    Location/Qualifiers
      1..1378
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1378;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 9
AX817901/c
LOCUS      1379 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 13 from patent US 6593093.
ACCESSION AR353663
VERSION    AR353663.1 GI:33759696
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1379)
AUTHORS   Uhl,J.R. and Cockerill,F.R.
TITLE      Detection of group a Streptococcus
JOURNAL    Patent: US 6593093-A 13 15-JUL-2003;
            Mayo Foundation for Medical Education and Research; Rochester, MN
FEATURES   source
            Location/Qualifiers
              1..1379
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 10
AR353666/c
LOCUS      1379 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 16 from patent US 6593093.
ACCESSION AR353666
VERSION    AR353666.1 GI:33759699
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1379)
AUTHORS   Uhl,J.R. and Cockerill,F.R.
TITLE      Detection of group a Streptococcus
JOURNAL    Patent: US 6593093-A 16 15-JUL-2003;
            Mayo Foundation for Medical Education and Research; Rochester, MN
FEATURES   source
            Location/Qualifiers
              1..1379
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 11
AX817894/c
LOCUS      1379 bp      DNA      linear      PAT 10-DEC-2003
DEFINITION Sequence 7 from Patent EPI338656.
ACCESSION AX817894
VERSION    AX817894.1 GI:39723080
KEYWORDS   .
SOURCE     Streptococcus sp. 'group A'

FEATURES
  source
    Location/Qualifiers
      1..1378
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1378;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 7
AX817901/c
LOCUS      1378 bp      DNA      linear      PAT 10-DEC-2003
DEFINITION Sequence 14 from Patent EPI338656.
ACCESSION AX817901
VERSION    AX817901.1 GI:39723087
KEYWORDS   .
SOURCE     Streptococcus sp. 'group A'
ORGANISM   Streptococcus sp. 'group A'
REFERENCE  1 (bases 1 to 1378)
AUTHORS   Uhl,J.R. and Cockerill Iii,F.R.
TITLE      Detection of group a Streptococcus
JOURNAL    Patent: EP 1338656-A 14 27-AUG-2003;
            MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES   source
            Location/Qualifiers
              1..1378
              /organism="Streptococcus sp. 'group A'"
              /mol_type="unassigned DNA"
              /db_xref="taxon:36470"
              /note="ptai sequence from isolate no. 10"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1378;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 8
AR353657/c
LOCUS      1379 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 7 from patent US 6593093.
ACCESSION AR353657
VERSION    AR353657.1 GI:33759690
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1379)
AUTHORS   Uhl,J.R. and Cockerill,F.R.
TITLE      Detection of group a Streptococcus
JOURNAL    Patent: US 6593093-A 7 15-JUL-2003;
            Mayo Foundation for Medical Education and Research; Rochester, MN
FEATURES   source
            Location/Qualifiers
              1..1379
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
  Query Match      100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTTATCTTCC 189
```

```

ORGANISM      Streptococcus sp. 'group A'
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE      1
AUTHORS        Uhl,J.R. and cockerill Iii,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: EP 1338656-A 7 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source       Location/Qualifiers
              1. .1379
                /organism="Streptococcus sp. 'group A'"
                /mol_type="unassigned DNA"
                /db_xref="taxon:36470"
                /note="ptseI sequence from isolate no. 6"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||||||||||||||
DB      209 TGCATGTATGGGTTATCTTCC 189
        |||||||||||||||||||
RESULT 12
AX817900/c      1379 bp DNA linear PAT 10-DEC-2003
LOCUS
DEFINITION      Sequence 13 from Patent EP1338656.
ACCESSION        AX817900
VERSION          AX817900.1 GI:39723086
KEYWORDS         Streptococcus sp. 'group A'
SOURCE           Streptococcus sp. 'group A'
ORGANISM         Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
                 Streptococcus.
REFERENCE      1
AUTHORS        Uhl,J.R. and cockerill Iii,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: EP 1338656-A 13 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source       Location/Qualifiers
              1. .1379
                /organism="Streptococcus sp. 'group A'"
                /mol_type="unassigned DNA"
                /db_xref="taxon:36470"
                /note="ptseI sequence from isolate no. 4"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||||||||||||||
DB      209 TGCATGTATGGGTTATCTTCC 189
        |||||||||||||||||||
RESULT 13
AX817903/c      1379 bp DNA linear PAT 10-DEC-2003
LOCUS
DEFINITION      Sequence 16 from Patent EP1338656.
ACCESSION        AX817903
VERSION          AX817903.1 GI:39723089
KEYWORDS         Streptococcus sp. 'group A'
SOURCE           Streptococcus sp. 'group A'
ORGANISM         Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
                 Streptococcus.
REFERENCE      1
AUTHORS        Uhl,J.R. and cockerill Iii,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: EP 1338656-A 16 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source       Location/Qualifiers
              1. .1379
                /organism="Streptococcus sp. 'group A'"
                /mol_type="unassigned DNA"
                /db_xref="taxon:36470"
                /note="ptseI sequence from isolate no. 4"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||||||||||||||
DB      209 TGCATGTATGGGTTATCTTCC 189
        |||||||||||||||||||
RESULT 14
AR353658/c      1384 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION      Sequence 8 from patent US 6593093.
ACCESSION        AR353658
VERSION          AR353658.1 GI:33759691
KEYWORDS         Unknown.
SOURCE           Unknown.
ORGANISM         Unclassified.
REFERENCE      1 (bases 1 to 1384)
AUTHORS        Uhl,J.R. and Cockerill,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: US 6593093-A 8 15-JUL-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES
  source       Location/Qualifiers
              1. .1384
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||||||||||||||
DB      210 TGCATGTATGGGTTATCTTCC 190
        |||||||||||||||||||
RESULT 15
AR353660/c      1384 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION      Sequence 10 from patent US 6593093.
ACCESSION        AR353660
VERSION          AR353660.1 GI:33759693
KEYWORDS         Unknown.
SOURCE           Unknown.
ORGANISM         Unclassified.
REFERENCE      1 (bases 1 to 1384)
AUTHORS        Uhl,J.R. and Cockerill,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: US 6593093-A 10 15-JUL-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES
  source       Location/Qualifiers
              1. .1384
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 209 TGCATGATGGGTTATCTTCC 189

RESULT 16
AX817895/c
LOCUS AX817895 1384 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 8 from Patent EP1338656.
ACCESSION AX817895
VERSION AX817895.1 GI:39723081
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl,J.R. and cockerill Iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 8 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    source
        1. .1384
        /organism="Streptococcus sp. 'group A'"
        /mol_type="unassigned DNA"
        /db_xref="taxon:36470"
        /note="ptseI sequence from isolate no. 5"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 17
AX817897/c
LOCUS AX817897 1384 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 10 from Patent EP1338656.
ACCESSION AX817897
VERSION AX817897.1 GI:39723083
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl,J.R. and cockerill Iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 10 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    source
        1. .1384
        /organism="Streptococcus sp. 'group A'"
        /mol_type="unassigned DNA"
        /db_xref="taxon:36470"
        /note="ptseI sequence from isolate no. 8"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 209 TGCATGATGGGTTATCTTCC 189

RESULT 18
AX817895/c
LOCUS AX817895 1385 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 9 from patent US 6593093.
ACCESSION AR353659
VERSION AR353659.1 GI:33759692
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1385)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: US 6593093-A 9 15-JUL-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES
    source
        1. .1385
        /organism="unknown"
        /mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1385;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 19
AX817896/c
LOCUS AX817896 1385 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 9 from Patent EP1338656.
ACCESSION AX817896
VERSION AX817896.1 GI:39723082
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl,J.R. and cockerill Iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 9 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    source
        1. .1385
        /organism="Streptococcus sp. 'group A'"
        /mol_type="unassigned DNA"
        /db_xref="taxon:36470"
        /note="ptseI sequence from isolate no. 7"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1385;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 20
AX817896/c
LOCUS AR353661 1390 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6593093.
ACCESSION AR353661
VERSION AR353661.1 GI:33759694
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1390)
AUTHORS Uhl,J.R. and Cockerill,F.R.

```

TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 11 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
Location/Qualifiers  
1. .1390  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 224 TGCATGTATGGGTTATCTTCC 204  
RESULT 21  
LOCUS AX817898/c 1390 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 11 from Patent EPI338656.  
ACCESSION AX817898  
VERSION AX817898.1 GI:39723084  
KEYWORDS  
SOURCE Streptococcus sp. 'group A'  
ORGANISM Streptococcus sp. 'group A',  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Uhl,J.R. and cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 11 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
Location/Qualifiers  
1. .1390  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptSI sequence from isolate no. 9"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 224 TGCATGTATGGGTTATCTTCC 204  
RESULT 22  
LOCUS AR353665/c 1393 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 15 from patent US 6593093.  
ACCESSION AR353665  
VERSION AR353665.1 GI:33759698  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1393)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 15 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
Location/Qualifiers  
1. .1393  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190  
RESULT 23  
LOCUS AX817902/c 1393 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 15 from Patent EPI338656.  
ACCESSION AX817902  
VERSION AX817902.1 GI:39723088  
KEYWORDS  
SOURCE Streptococcus sp. 'group A'  
ORGANISM Streptococcus sp. 'group A',  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Uhl,J.R. and cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 15 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
Location/Qualifiers  
1. .1393  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptSI sequence from isolate no. 3"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190  
RESULT 24  
LOCUS AR353667/c 1447 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 17 from patent US 6593093.  
ACCESSION AR353667  
VERSION AR353667.1 GI:33759700  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1447)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 17 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
Location/Qualifiers  
1. .1447  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190  
RESULT 25  
LOCUS AX817904/c 1447 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 17 from patent US 6593093.  
ACCESSION AR353667  
VERSION AR353667.1 GI:33759700  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1447)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 17 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
Location/Qualifiers  
1. .1447  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190

```

LOCUS      AX817904      1447 bp      DNA      linear      PAT 10-DRC-2003
DEFINITION Sequence 17 from Patent EP1338656.
ACCESSION  AX817904
VERSION     AX817904.1  GI:39723090
SOURCE      Streptococcus sp. 'group A'
ORGANISM    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
AUTHORS     Uhl,J.R. and cockerill Iii,F.R.
TITLE       Detection of group A Streptococcus
JOURNAL     Patent: EP 1338656-A 17 27-AUG-2003;
            MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES    Location/Qualifiers
            source
            1..1447
            /organism="Streptococcus sp. 'group A'"
            /mol_type="unassigned DNA"
            /db_xref="taxon:36470"
            /note="ptsi sequence from isolate no. 11"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1447;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTATCTTCC 21
        |||||||
Db      210 TGCATGTATGGGTATCTTCC 190

RESULT 26
CO647508/c
LOCUS      CO647508      1731 bp      DNA      linear      PAT 02-FEB-2004
DEFINITION Sequence 4465 from Patent WO0234771.
ACCESSION  CO647508
VERSION     CO647508.1  GI:41683480
KEYWORDS    Streptococcus agalactiae
ORGANISM    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
AUTHORS     Telford,J., Massignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
            and Tettelin,H.
TITLE       Nucleic acids and proteins from streptococcus groups a & b
JOURNAL     Patent: WO 0234771-A 4465 02-MAY-2002;
            Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES    Location/Qualifiers
            source
            1..1731
            /organism="Streptococcus agalactiae"
            /mol_type="unassigned DNA"
            /db_xref="taxon:1311"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTATCTTCC 21
        |||||||
Db      377 TGCATGTATGGGTATCTTCC 357

RESULT 27
CO647510/c
LOCUS      CO647510      1731 bp      DNA      linear      PAT 02-FEB-2004
DEFINITION Sequence 4467 from Patent WO0234771.
ACCESSION  CO647510
VERSION     CO647510.1  GI:41683481
KEYWORDS    Streptococcus pyogenes
ORGANISM    Streptococcus pyogenes

```

---

```

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE   1
AUTHORS     Telford,J., Massignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
            and Tettelin,H.
TITLE       Nucleic acids and proteins from streptococcus groups a & b
JOURNAL     Patent: WO 0234771-A 4467 02-MAY-2002;
            Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES    Location/Qualifiers
            source
            1..1731
            /organism="Streptococcus pyogenes"
            /mol_type="unassigned DNA"
            /db_xref="taxon:1314"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTATCTTCC 21
        |||||||
Db      377 TGCATGTATGGGTATCTTCC 357

RESULT 28
AX608629/c
LOCUS      AX608629      1734 bp      DNA      linear      PAT 17-FEB-2003
DEFINITION Sequence 6558 from Patent WO02092818.
ACCESSION  AX608629
VERSION     AX608629.1  GI:28404206
KEYWORDS    Streptococcus agalactiae
ORGANISM    Streptococcus agalactiae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
AUTHORS     Glaser,P., Rueniok,C., Chevalier,F., Prangeul,L., Lalloui,L.,
            Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
            Kunst,F.
TITLE       Streptococcus agalactiae genome sequence, use for developing
            vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL     Patent: WO 02092818-A 6558 21-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES    Location/Qualifiers
            source
            1..1734
            /organism="Streptococcus agalactiae"
            /mol_type="unassigned DNA"
            /db_xref="taxon:1311"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGCATGTATGGGTATCTTCC 21
        |||||||
Db      377 TGCATGTATGGGTATCTTCC 357

RESULT 29
AR353656/c
LOCUS      AR353656      1803 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 5 from patent US 6593093.
ACCESSION  AR353656
VERSION     AR353656.1  GI:33759689
KEYWORDS    Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
            1 (bases 1 to 1803)
REFERENCE   1
AUTHORS     Uhl,J.R. and Cockerill,F.R.
TITLE       Detection of group A Streptococcus

```

JOURNAL	Patent: US 6593093-A 5 15-JUL-2003;	TITLE	Direct Submission
FEATURES	Mayo Foundation for Medical Education and Research; Rochester, MN	JOURNAL	Submitted (24-MAY-1999) Narito Aasunuma, Meiji University,
source	Location/Qualifiers		Department of Agriculture; Higashimita, Tama-ku, Kawasaki, Kanagawa
	1. .1803		214-8571, Japan (E-mail: aasunuma@sc.meiji.ac.jp,
	/organism="unknown"		Tel.: +81-44-934-7825 (ex.7825), Fax: +81-44-934-7825)
	/mol_type="genomic DNA"		On Jan 20, 2000 this sequence version replaced gi:5706360.
ORIGIN		FEATURES	Location/Qualifiers
	Query Match 100.0%; Score 21; DB 6; Length 1803;	source	1. .3797
	Best Local Similarity 100.0%; Pred. No. 10;		/organism="Streptococcus bovis"
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		/mol_type="genomic DNA"
			/strain="JBI"
Qy	1 TGCATGTATGGTTATCTTCC 21	gene	/db_xref="taxon:1315"
			1779..2042
			/gene="ptsH"
			1779..2042
			/gene="ptsH"
Db	377 TGCATGTATGGTTATCTTCC 357	CDS	/codon_start=1
			/transl_table=1
			/product="histidine containing protein"
RESULT 30			/protein_id="BAA78048.1"
AX817892/c	AX817892 1803 bp DNA linear PAT 10-DEC-2003		/db_xref="GI:4958915"
DEFINITION	Sequence 5 from Patent EP1338656.		/translation="MASKDPHIVAFGTGTHARPATLLVQTASKFASDITLDYKGAIVNL
ACCESSION	AX817892		KSIMGVMSLGVGGQADVTISAGADADDALAAIBETMTKEGLA"
VERSION	AX817892.1 GI:39723079	gene	2047..3780
KEYWORDS			/gene="ptsI"
SOURCE	Streptococcus sp. 'group A'	CDS	2047..3780
ORGANISM	Streptococcus sp. 'group A',		/gene="ptsI"
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		/codon_start=1
	Streptococcus.		/transl_table=1
REFERENCE	1		/product="phosphoenolpyruvate-protein phosphotransferase"
AUTHORS	Uhl, J.R. and cockerill Iii, F.R.		/protein_id="BAA78049.1"
TITLE	Detection of group A Streptococcus		/db_xref="GI:4958916"
JOURNAL	Patent: EP 1338656-A 5 27-AUG-2003;		/translation="MTEMLKGIASDGVAVAKAYLLVQPDLSFETVTVEDTSAEARL
FEATURES	MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)		DAALKASQDELIIIEKAVETIGEEAAVFDALHMLVADPEMISQIKETIRAKQTNAE
source	Location/Qualifiers		AGLKEVTDMTIFIEGEMEDNPMQERAAIDRVAKRVLAHLIGAKLPNPATIDESSIV
	1. .1803		IAHDLTPSDTQALNKFVKAFVNIIGRTSHSAIMARTLEAAVLGTNDITSRVKQDG
	/organism="Streptococcus sp. 'group A'"		IVAVNGITGEVIINPTDQEAEFKAAGEAYAKQAEWALLDKAKTVTADGKHFELEAAN
	/mol_type="unassigned DNA"		IGTPADVEGVNANGAEAVGLYRTFLYHDSQDFPEDQYENAYKAVLGMGKGPVVR
	/db_xref="taxon:36470"		TMDIGGDKELPDLPKEMNPFGLRALRISISSETGNAMFRTIIRALLRASVHQGLRI
	/note="ptsI sequence from Oklahoma University M1 strain"		MFMVALLKEFAAKAIPEDEKALKAEGVAVSDDIQVIGIMIEIPAAHMLADQFAKEV
ORIGIN		ORIGIN	MAGDQKAVPLLVEMLGDFMSATSILRTRSLMKKLDIAKMQEYANRALTEGSTMEEV
	Query Match 100.0%; Score 21; DB 6; Length 1803;		LELSKEYNVND"
	Best Local Similarity 100.0%; Pred. No. 10;	Query Match	100.0%; Score 21; DB 1; Length 3797;
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Best Local Similarity	100.0%; Pred. No. 9.1;
		Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 TGCATGTATGGTTATCTTCC 21		
Db	377 TGCATGTATGGTTATCTTCC 357	Qy	1 TGCATGTATGGTTATCTTCC 21
RESULT 31		Db	2423 TGCATGTATGGTTATCTTCC 2403
AB027569/c	AB027569 3797 bp DNA linear BCT 23-JUL-2003		
LOCUS	Streptococcus bovis ptsH, ptsI genes for histidine containing		
DEFINITION	protein, phosphoenolpyruvate-protein phosphotransferase, complete		
	cds.		
ACCESSION	AB027569		
VERSION	AB027569.3 GI:6723275	RESULT 32	
KEYWORDS	phosphoenolpyruvate-protein phosphotransferase; histidine	AB006575	
	containing protein.	LOCUS	
SOURCE	Streptococcus bovis	DEFINITION	Streptococcus pyogenes M1 GAS, section 104 of the complete
ORGANISM	Streptococcus bovis	ACCESSION	AB006575 AE004092
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	VERSION	AB006575.1 GI:13622480
	Streptococcus.	KEYWORDS	
REFERENCE	1	SOURCE	Streptococcus pyogenes M1 GAS
AUTHORS	Asanuma, N. and Hino, T.	ORGANISM	Streptococcus pyogenes M1 GAS
TITLE	Molecular characterization of Hpr and related enzymes, and		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
	regulation of Hpr phosphorylation in the ruminal bacterium		Streptococcus.
JOURNAL	Streptococcus bovis	REFERENCE	1 (bases 1 to 10430)
PUBMED	Arch. Microbiol. 179 (3), 205-213 (2003)	AUTHORS	Perretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
REFERENCE	12610726		Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
AUTHORS	Asanuma, N. and Hino, T.		Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
			Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
			Complete genome sequence of an M1 strain of Streptococcus pyogenes
			Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

PUBMED REFERENCE AUTHORS	11296296 2 (bases 1 to 10430) Ferretti,J., Mcshan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najat,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E. Direct Submission
TITLE JOURNAL	Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104, USA Location/Qualifiers
FEATURES source	1. .10430 /organism="Streptococcus pyogenes M1 GAS" /mol_type="genomic DNA" /strain="SF370" /serotype="M1" /db_xref="taxon:160490" complement(78. .1505) /gene="gapN" /locus_tag="SPy1371" complement(78. .1505) /gene="gapN" /locus_tag="SPy1371" /EC_number="1.2.1.9" /note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of <i>S. pyogenes</i> . Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of <i>Streptococcus pyogenes</i> proteins. Unpublished data.; Best Blastp hit = dbj BA05956.1  (AP001514) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase [Bacillus halodurans]" /codon_start=1 /transl_table=11 /product="putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase" /protein_id="AAK34198.1" /db_xref="GI:13622481"
gene	/note="Best Blastp hit = sp P30299 P71_STRSL PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME 1) >gi 322087 pir  JCI1375 phosphotransferase system enzyme I (EC 2.7.3.9) - Streptococcus salivarius >gi 153615 gb AAK26873.1  (M81756) phosphoenolpyruvate:sugar phosphotransferase system enzyme I I Streptococcus salivarius] >gi 1583334 prf  21203818 ptai gene [Streptococcus salivarius]" /codon_start=1 /transl_table=11 /product="putative phosphoenolpyruvate:sugar phosphotransferase system enzyme I" /protein_id="AAK34199.1" /db_xref="GI:13622482"
CDS	/translation="MTMLKGIASDGVAVAKAYLLVQPOLSFETVTDVADTNAEEARL DVALQAQDELSVIRENAVSLGEAAAVFDHMLVIADPEMISQVKETIRAKOTNAE TGLKEVDMFTITEGHEDPYMQERAADIDVAKRVLAHLGLVKNPNTPEESIV
gene	/note="Best Blastp hit = sp Q48708 NRDH LACLC GLUTAREDOXIN-LIKE PROTEIN NRDH >gi 1222680 emb CAA63372.1  (X92690) glutaredoxin-like protein [Lactococcus lactis]" /codon_start=1 /transl_table=11 /product="putative glutaredoxin" /protein_id="AAK34201.1" /db_xref="GI:13622484"
CDS	/translation="MTIVYSKNNCKMCKKFKLEQHGYNFQENIDEHPEKVDYVKS LGFTSAPVIEADNLVPSGFQPAKLKELI" 4322. .6481 /gene="nrde.2" /locus_tag="SPy1375" 4322. .6481 /gene="nrde.2" /EC_number="1.17.4.1" /note="Best Blastp hit = gb AAD41036.1 AF112535_3 (AF112535) ribonucleotide reductase alpha-chain [Corynebacterium glutamicum]" /codon_start=1 /transl_table=11 /product="putative ribonucleotide reductase alpha-chain" /protein_id="AAK34202.1" /db_xref="GI:13622485"
gene	/translation="MSLKLDGDISYERLNNENRPNVNGKIPLHKDKALKAFSAENVL PNTWSTFSTKELEYLISNDYIESAFIQKRPPIFELDSIIXSENFKFSFMAVKF YQYVALTKNDGHEYNLEDRVFNALYFADGGEDLAKDLAVEMINORYOPATPSFLN AGRSRGELVSCFLIOVTDMSINSIGRSINSLQLSRIGGGVITLSNLRBAGAPIKGY AGASGVVPVWMLFEDSFYSNQLGQRGAGVYVLYNVFHPDIIIAFLSTRKENADEKVR
CDS	/translation="MSLKLDGDISYERLNNENRPNVNGKIPLHKDKALKAFSAENVL PNTWSTFSTKELEYLISNDYIESAFIQKRPPIFELDSIIXSENFKFSFMAVKF YQYVALTKNDGHEYNLEDRVFNALYFADGGEDLAKDLAVEMINORYOPATPSFLN AGRSRGELVSCFLIOVTDMSINSIGRSINSLQLSRIGGGVITLSNLRBAGAPIKGY AGASGVVPVWMLFEDSFYSNQLGQRGAGVYVLYNVFHPDIIIAFLSTRKENADEKVR
gene	/note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of <i>S. pyogenes</i> . Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of <i>Streptococcus pyogenes</i> proteins. Unpublished data.; Best Blastp hit = sp P34366 P7HP_STRSL PHOSPHOCARRIER PROTEIN HPR (HISTIDINE-CONTAINING PROTEIN) >gi 480991 pit  837585 phosphotransferase system phosphohistidine-containing protein - Streptococcus salivarius >gi 406779 emb CAA78923.1  (Z17217) HPR [Streptococcus salivarius] >gi 739971 prf  2004268A His containing phosphocARRIER protein [Streptococcus salivarius] >gi 1583333 prf  2120381A ptai gene [Streptococcus salivarius]" /codon_start=1 /transl_table=11 /product="putative phosphotransferase system phosphohistidine-containing protein" /protein_id="AAK34200.1" /db_xref="GI:13622483"
gene	/translation="WASKDFHVAETGIHARPATLLVQTASKFASDITLDYKGAENVL KSLMGVMSLGVGGADVTIISAEAGDAEDAIAIEETMTKEGLA" 4084. .4302 /gene="nrDH" /locus_tag="SPy1374" 4084. .4302 /gene="nrDH" /locus_tag="SPy1374" /note="Best Blastp hit = sp Q48708 NRDH LACLC GLUTAREDOXIN-LIKE PROTEIN NRDH >gi 1222680 emb CAA63372.1  (X92690) glutaredoxin-like protein [Lactococcus lactis]" /codon_start=1 /transl_table=11 /product="putative glutaredoxin" /protein_id="AAK34201.1" /db_xref="GI:13622484"
CDS	/translation="MTIVYSKNNCKMCKKFKLEQHGYNFQENIDEHPEKVDYVKS LGFTSAPVIEADNLVPSGFQPAKLKELI" 4322. .6481 /gene="nrde.2" /locus_tag="SPy1375" 4322. .6481 /gene="nrde.2" /EC_number="1.17.4.1" /note="Best Blastp hit = gb AAD41036.1 AF112535_3 (AF112535) ribonucleotide reductase alpha-chain [Corynebacterium glutamicum]" /codon_start=1 /transl_table=11 /product="putative ribonucleotide reductase alpha-chain" /protein_id="AAK34202.1" /db_xref="GI:13622485"
gene	/translation="MSLKLDGDISYERLNNENRPNVNGKIPLHKDKALKAFSAENVL PNTWSTFSTKELEYLISNDYIESAFIQKRPPIFELDSIIXSENFKFSFMAVKF YQYVALTKNDGHEYNLEDRVFNALYFADGGEDLAKDLAVEMINORYOPATPSFLN AGRSRGELVSCFLIOVTDMSINSIGRSINSLQLSRIGGGVITLSNLRBAGAPIKGY AGASGVVPVWMLFEDSFYSNQLGQRGAGVYVLYNVFHPDIIIAFLSTRKENADEKVR
CDS	/translation="MSLKLDGDISYERLNNENRPNVNGKIPLHKDKALKAFSAENVL PNTWSTFSTKELEYLISNDYIESAFIQKRPPIFELDSIIXSENFKFSFMAVKF YQYVALTKNDGHEYNLEDRVFNALYFADGGEDLAKDLAVEMINORYOPATPSFLN AGRSRGELVSCFLIOVTDMSINSIGRSINSLQLSRIGGGVITLSNLRBAGAPIKGY AGASGVVPVWMLFEDSFYSNQLGQRGAGVYVLYNVFHPDIIIAFLSTRKENADEKVR

VKTLSLGIITVPDKFYELARKNEDMYLFSPYNVEKEYGIPENLYNDITNMVDELVANPKI  
TTKTKARLETEISKLOQESVPIIINIDTANKANPIDGKILMNSCLSEILQVOTPS  
LINDQAEFVEMGTDISCNLGSNINLMNTSPFGRSIKTWTALFTVTDSSIEAVPT  
IKHNSQAHTFGLGANGSLYLAQHIIYEGSPESIEFTDIYFMLLNWTLVESNNIAR  
ERQTTVPFENSKYANGSYFDKYVTHGFYFKSDLVKDLFKDHFIFQASDWEALRDVQ  
KDGLYHQNRLAVAPNGISYINDCSASIHPIQTORIEERQEKIKIYYIPANGLSDTDI  
PYTSAIDMDMRKVIDVYAAATEHVDDQGLSLTLFLRSELPMELYEWKTSQKTTDRDLS  
ILRNVAFNKGIKSIYYIRFTDDGEEVGANQCESCVI"  
6814. .7773  
/gene="nrdf.2"  
/locus\_tag="SPy1378"  
6814. .7773  
/gene="nrdf.2"  
/locus\_tag="SPy1378"  
/note="Best Blastp hit = emb|CAA70766.1| (Y09572)  
ribonucleotide reductase subunit R2P [Corynebacterium  
ammoniaenes]  
/codon\_start=1  
/transl\_table=11  
/product="ribonucleotide diphosphate reductase small  
subunit"  
/protein\_id="AAK34203.1"  
/db\_xref="GI:13622486"  
/translation="MTYYEAINWNIEVIDIKSTWEKLTQEOWLDTRIPLNSDLDDW  
RKLQEKDLVGKFGGLTLLDTMQSETGVEAIRADVRTPHEEAVLNNTQFMSVHAK  
SYSSIFSTLNTKKEIEEIEFWTNNEFLQEKARIINDIYANGDALQKKVASTYSLTEL  
FVSGFTPLIYGNKKLANVASIIKLIINDESVDHYIYGKQGLGNEFLPEDEQENFR  
DMYDLYQLYNEERKTYKLYDGVGWTEEVMTFLRYNANKALMNLGQDLPLFPDAND  
VNPIVMNGISTGTSNHDFFSQVGVNGYLLGSVEAMSDDDYNYGL"  
7748. .9061  
/locus\_tag="SPy1379"  
7748. .9061  
/locus\_tag="SPy1379"  
/note="Best Blastp hit = gb|AAF41386.1| (AE002449)  
chloride channel protein-related protein [Neisseria  
meningitidis Mc58]  
/codon\_start=1  
Query Match 100.0%; Score 21; DB 1; Length 10430;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 3047 TGCATGTATGGGTATCTTCC 3067  
RESULT 33  
AE010058  
LOCUS 11992 bp DNA linear BCT 03-APR-2002  
DEFINITION Streptococcus pyogenes strain MGAS8232, section 106 of 173 of the  
complete genome.  
ACCESSION AE010058 AE009949  
VERSION AE010058.1 GI:19748529  
KEYWORDS  
SOURCE  
ORGANISM Streptococcus pyogenes MGAS8232  
Streptococcus pyogenes MGAS8232  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 11992)  
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,  
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.  
TITLE Genome sequence and comparative microarray analysis of serotype M18  
group A Streptococcus strains associated with acute rheumatic fever  
outbreaks  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
PUBMED 11917108  
REFERENCE 2 (bases 1 to 11992)  
AUTHORS Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,

Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Musser,J.M.  
Direct Submission  
Submitted (31-JAN-2002) Laboratory of Human Bacterial  
Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St.,  
Hamilton, MT 59840, USA  
FEATURES  
source  
1. .11992  
/organism="Streptococcus pyogenes MGAS8232"  
/mol\_type="genomic DNA"  
/strain="MGAS8232"  
/db\_xref="taxon:186103"  
132. .1073  
/gene="bira"  
132. .1073  
/note="synonym: spyM18\_1374"  
/gene="bira"  
/note="best blastp match gb|AAK34189.1| (AE006574)  
putative biotin operon repressor [Streptococcus pyogenes  
M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAL97969.1"  
/db\_xref="GI:19748528"  
/translation="MKTSEKIVQLLSQTDQDFVSGEYLADQLSISRTSVWKSIXLSLENO  
GIQDSLAKHKYRVWQDILLPKTISQGLGMPVTVTPHSQSTQLDKOGIEAHNSAPR  
LYLAPQSAAGRLDRQFSASTGGIYNSWYLNKENVPDMPPTVMVASSIVKAISR  
LTGDTIEIKVNDIYLGHNKAVAGILTEAITSVETGLITDVIIGVGNFDFPDAIA  
QKASLFTKEPTITRNLDIIDIWKLFLSIPVKDHVKYKSEKSLVNLKQVTFIENSQEK  
RAAIDLTDDQGLHVQFDNGLQTLRSGEISLSWEN"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
/note="best blastp match gb|AAK34190.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97970.1"  
/db\_xref="GI:19748529"  
/translation="MPLPTFYLVTFHFHSTVFCYKNAMTIRQFLFMAFVCAFETYFFN  
DLLSGNYLFALFWGLLFRDLRRVHTINQLTKTLKAANSPPKKD"  
complement (1346. .3016)  
/gene="dnaX"  
/note="synonym: spyM18\_1376"  
complement (1346. .3016)  
/gene="dnaX"  
/note="best blastp match gb|AAF98348.1| (AF280764) DNA  
polymerase III gamma/tau subunits [Streptococcus  
pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAL97971.1"  
/db\_xref="GI:19748530"  
/translation="WYQALYKRSQTFEDVMVQSVISTTLKQAVESGKISHAYLPSPG  
PRGTGKTSAAKIFAKMCPNQVDEPCNCDICRDTNGSLSEVDIEDIASQNSGVDE  
IRDIRKSTYAPSRATYKYIIDEVHMLSTGAFNALLKTBEETPNVVFILATTELHK  
IPATILSRVQRFQKAIQKRAIRHAWLDKEGIAYEVDALNLTARAEGRMDALLS  
ILDQALSLPDNQVTIAAEBITGSGISILALGVYRVYVQSQATQALAALETIYDSGK  
SMSRPATLLYLRDLVVKAGDNQROSAVFDNLSLIDRIQOMITVWTHLPEIK  
KGTPTPIYMMTTLQAKQKILSQANLSGLISEIETLKNELAQKQOLQSORPDK  
SLAKSDTKPTKTTSTYRDTYLLKIMEETVRNSQSRQYLRDLAKNAWEILDNISAQD  
RALLMGSPFLANSENALAFEAFAFNAEQVMNRNLDNMFNGINMSKAAGFSNLAAP  
RTDFQHIRKEFAQMKSQKSDVQBEQVEALDIPEGDFDILLDKINTIDD"  
complement (3016. .3513)  
/gene="spyM18\_1377"  
complement (3016. .3513)  
/gene="spyM18\_1377"  
/note="best blastp match gb|AAK34192.1| (AE006574)  
conserved hypothetical protein [Streptococcus pyogenes M1  
GAS]"

gene	/codon_start=1	CDS	/gene="deadD2"	Query Match 100.0%; Score 21; DB 1; Length 11992; Best Local Similarity 100.0%; Pred. No. 7.5; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAL97972.1"				
	/db_xref="GI:19748531"				
	/translation="MNKSKIEYOQLMIAQAKELFANESNALANISNALINMTLPN				
	SVYTFGLFDGQELILGPGVSCVHIKLGKVGCGESAQRRTIIINDVKQHANYIS				
	CDAAAMSEIVPMVKEGHILGVLDDSSILVADYDEVQBYLETFFVDLLLEKTTFTFNM				
	FGYKN"				
	3658..4467				
CDS	/gene="spyM18_1378"	gene	/product="putative deacetylase"		
	/protein_id="AAL97977.1"				
	/db_xref="GI:19748536"				
	/translation="MNKLVILVGLLSILMLSLAIVFIRNWKLNEDSORIVLAEKKQN				
	TSDLVIKAVKHKKDDQVYFSPKQADDFVDNLPLVSLYKKKSDRELILVKKPLQ				
	SHLSRSVNTLTISKIVYQKPFHLAKSEKVIKSTYHVDNLKPFQVKDLVSGHLERIQ				
	BEVEKKYNAGNSDKYNGKESNLSDGFEVKSNGLIFDKLLTIPLTFLDFVINPD				
	FIANSDRAAYDNYRTYKEQHPKKLVALTFDDQDPDPTTTPQVLDILAKYQAGTFFPMIG				
	SKVYVNEMLTKREVSADAGIEAHNHTWDPNLNLNSVSEIQHVNMTNQATEKACGKKPR				
	YLRPPYGATNAVTVQSSSLGTQMLWTVDTRDWHNSTDGMTNWKQLQFGGVVLMDHI				
HQTINALPTVMEYLKAGYECCVTVSELYAHQ"					
gene	complement (8153..9580)	CDS	/gene="gapN"		
	/note="synonym: spyM18_1383"				
	complement (8153..9580)				
	/gene="gapN"				
	/note="best blastp match gb AAK34198.1  (AE006575)"				
	putative NADP-dependent glyceraldehyde-3-phosphate				
	dehydrogenase [Streptococcus pyogenes M1 GAS]"				
	/codon_start=1				
	/transl_table=11				
	/product="putative NADP-dependent				
Qy	1 TGCATGTATGGGTATCTTCC 21	Db	11122 TGCATGTATGGGTATCTTCC 11142		
RESULT 34	AE014229/c	LOCUS	AE014229		
	21781 bp				
	DNA				
	linear				
	BCT 23-SEP-2002				
	Streptococcus agalactiae 2603V/R section 39 of 100 of the complete genome.				
	AE014229 AE009948				
	AE014229.1 GI:22533831				
	Streptococcus agalactiae 2603V/R				
	Streptococcus agalactiae 2603V/R				
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.	AUTHORS	1 (bases 1 to 21781)		
	Tettelin, H., Maignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Dougherty, S.C., DeBoy, R.T., Durkin, S., Kolonay, J.F., Umayan, L.A., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Gill, J., Scarcelli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G., Mariani, M., Vegni, F., Malone, D., Rinaldo, D., Rappuoli, R., Telford, J.L., Kasper, D.L., Grandi, G. and Fraser, C.M.				
	Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae				
	Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)				
	12200547				
	2 (bases 1 to 21781)				
	Tettelin, H., Maignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Dougherty, S.C., DeBoy, R.T., Durkin, S., Kolonay, J.F., Umayan, L.A., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Gill, J.,				
	JOURNAL		PUBMED	REFERENCE	AUTHORS

AE014229 21781 bp DNA linear BCT 23-SEP-2002  
Streptococcus agalactiae 2603V/R section 39 of 100 of the complete genome.  
AE014229 AE009948  
AE014229.1 GI:22533831  
Streptococcus agalactiae 2603V/R  
Streptococcus agalactiae 2603V/R  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
1 (bases 1 to 21781)  
Tettelin,H., Massignani,V., Cieslewicz,M.J., Eisen,J.A.,  
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,  
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,  
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,  
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,  
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,  
Mariani,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,  
Marini,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,  
Telford,J.L., Kasper,D.L., Grandi,G. and Frazer,C.M.  
Complete genome sequence and comparative genomic analysis of an  
emerging human pathogen, serotype V Streptococcus agalactiae  
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)  
1220547  
2 (bases 1 to 21781)  
Tettelin,H., Massignani,V., Cieslewicz,M.J., Eisen,J.A.,  
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,  
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,  
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,  
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,  
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,





```

/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAM99708.1"
/db_xref="GI:225333840"
/translation="NASKDFHVAETGHIARPATLLVQTASKFASDITLDYKGAVALN
KSMGWSLGVGGADVTISAEGADADDAIAAIEETMTKEGLA"
6858..8591
/genes="psi1"
/notes="synonym: SAG0822"
6858..8591
/genes="psi1"
/notes="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAM99709.1"
/db_xref="GI:225333841"
/translation="WTEMLKGIAASDGVAVAKAYLLVQDLSFETVTVEDTNAEARL
DVALQSDLSVIREKAVESGEAAAVFDHLMLVLSDFEMINQIKETIRAKQVNAE
TGLKVTDFMTIFEGEDNPYQERADIRDVAKRVLAHLGLVLPNPATINEESIV
IAHDLTPSDAQLNKQVFAFVNTNIGRTSHGAIWARTLEIAAIVLGNITERVQDQ
Query Match 100.0%; Score 21; DB 1; Length 21781;
Best Local Similarity 100.0%; Pred. No. 6.8; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTCCTTC 21
Db 7234 TGCATGTATGGTTCCTTC 7214

RESULT 35
AX602188/c
LOCUS 29072 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 117 from Patent WO2092818.
ACCESSION AX602188
VERSION AX602188.1 GI:28402057
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 117 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source 1..29072
location/Qualifiers
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 29072;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTCCTTC 21
Db 12272 TGCATGTATGGTTCCTTC 12252

RESULT 36
AE014157
LOCUS 51454 bp DNA linear BCT 19-JUL-2002
DEFINITION Streptococcus pyogenes MGAS315, section 22 of 37 of the complete
```

```

genome.
AE014157 AE014074
AE014157.1 GI:21904768
Streptococcus pyogenes MGAS315
Streptococcus pyogenes MGAS315
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Genome sequence of a serotype M3 strain of group A Streptococcus:
Phage-encoded toxins, the high-virulence phenotype, and clone
emergence
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
2 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Direct Submission
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
1..51454
Location/Qualifiers
/organism="Streptococcus pyogenes MGAS315"
/mol_type="genomic DNA"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/notes="group: A"
complement(150..1139)
/genes="grb"
/notes="synonym: SpvM3_1032"
complement(150..1139)
/genes="grb"
/notes="best blastp hit: gb|AAD26342.1|AF124403.1
(AE014003) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein G-related alpha 2M-bindingprotein"
/protein_id="AAM79639.1"
/db_xref="GI:21904769"
/translation="MGKEIKVKYFLRRSAFGLVAVASVLSVSTGSAVDSPIRQPIRII
PNGRITLNLGNAPKALNRNEERAIDELKQAIEDKEATTAEASSDALEALADQDA
DALQSEAAVQSDTAASDALEALADQDALQSEAAVVKADTAASDALEALADQDALQ
VQSEASVVKADTAASDALEALADQDALQSEAAVVKADTAASDALEALADQDALQ
SEAAVVKADTAASDALEALADQDALQSEAAVVKADTAASDALEALADQDALQ
TKDTPVVKKEQRQNVLTPTTGEESNPFPTAALAMVSTGALVVSCKKEN"
complement(1496..2710)
/genes="murZ"
/notes="synonym: SpvM3_1033"
complement(1496..2710)
/genes="murZ"
/notes="best blastp hit: gb|AAL97966.1| (AE010057) putative
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAM79640.1"
/db_xref="GI:21904770"
/translation="MAVSGAKNSVVALIPAILADDDIVLDGVPDISDVDSLIEMEL
MGATVNYHGDLEIDPRGQDIPMPYKINSLSRAVYFYGSLIGRGOAVVGLPGGCD
LGRPIDLHLKAFAMGVESVEGENNNLSGTQKJHGAHIYMDTVSVGATINTVAA
TKAQGKTVIENAAAREPEIIDVATLNNMGAHIRKAGDTIITIGVQKLHTRKIQVDP
RIEAGTYIALAAAIKGVKRTVNLVYEHLESFTAKLEMGVVRMTVEDDAIEVVKQESL
ATIKTSPYGPATDQQPLTPLLKADGRGTIIDTYEKRIINHVPPELMRMGADISVI
GGQIVQGPSRLTGAQVKATDLRAGAALVETAGLMAEGKTEITNIEFLRGYASIAKL
```

gene  
complement (2929.. .4125)  
/gene="metK"  
/note="synonym: SpyM3\_1034"  
complement (2929.. .4125)  
/gene="metK"  
/note="best blastp hit: gb|AAL97967.1| (AE010057)  
S-adenosylmethionine synthetase [Streptococcus pyogenes MGAS8232]"  
/codon\_start=1  
/transl\_table=11  
/product="putative S-adenosylmethionine synthetase"  
/protein\_id="AAM79641.1"  
/db\_xref="GI:21904771"  
/translation="MSERKLFITSESVSEGHDPKADQISDAILDAILAEDPRAHVAEE  
TCVYTSVHFGEISTTAYIDINRVVRDTIAIGYTEAFSEYFSAESVGHPSLVQSG  
DIAQVNEALRESRGTDLLSHGADQGMFCFAINTEPELMPPIPSLSHQLVRLLA  
ELRKGEISYLRAPDAKSQVTVYDEHDKPVRVDTVVISQHDPEATNDQIRVDIEKV  
IKAVIPADYLDQDTKFFINPTGRFVTGGPQSGSLTGRIIVDTYGYSRHGGGAFSG  
KQATKVDRSASVAARYIAKLVNAGLATKAEVLQAYAGVAQPVSVRVDTFGTSTVPE  
AVLEAAVRQVFDLRPAGIIOQLDLKRPYIKQTAAYGHMGRITDIDLFWHLNKVDALVE  
AVKTVL"  
complement (4662.. .7040)  
/gene="inlA"  
/note="synonym: SpyM3\_1035"  
complement (4662.. .7040)  
/gene="inlA"  
/note="best blastp hit: gb|AAK34188.1| (AE006574) putative  
internalin A precursor [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative internalin A precursor"  
/protein\_id="AAM79642.1"  
/db\_xref="GI:21904772"  
/translation="MKTKKVIILVGLLSQLTLACQSRGNGTYPITKTKQRKGWTS  
NKIPTKSKKNTKTHGVAGVDFPTDDGFIITKDSKILSKTDQGVVDHGDGSHFIF  
YADLKGPEYILPKGASLAKPAVAQASQSTSKVADPHHYEENPADIVAEDALGY  
TVRHDDHFYHILKSLSSGQTQAKAQVATRLPQTSPPSVSTATANGIPGLHFTSDGQ  
FNGQGVGTQKNIIVDHDGHLHPSIFADLRQGGWAHVADQYDPKAKKPAETHHTP  
ELSERKEYQEKLAYLAELKIDPSTIKRVEVTDQGLGLEYPHHDHAHVLMLSDEIG  
KDIPOPHATEHARELEKHKFGMDTLRALGFDEVIDIVRTHDAPTPPFSNEKDPNM  
KQWLTAVIKLIDGSRKDPQRLQKLSLPLETLTGIGFTPIKDIPVLPKPKLKQLMT  
KTGVDTYRFLDNWQLEGIDISQNLKDIISFKYKNLTVAANGIDIESLGOLP  
NULFVLVSNKISDLSPLASQLQELHIDNNQITDLSVPVSHKESITVVDLGRNADVD  
LATLPQAPLKTMDNTDKVSHLDPLFKNNPNLSLSINRAQSLLEGIEASSVIVRVEA  
EGNQIKSLVKDKQSGSLTFDTNQGLTLEGVNNFTALDILSVKNQITNVLNLSKPN  
KVTNIDIGHNNIPLADLNEQHIPEAIKAPFPAYVEGSMVGNGTAEKKAAMATNAK  
ESAOEASESHDYNHNHAYEDEGHAHEHRDKDDHDHEHEDENEAKDEQNHAH"  
7244.. .8185  
/gene="birA"  
/note="synonym: SpyM3\_1036"  
7244.. .8185  
/gene="birA"  
/note="best blastp hit: gb|AAK34189.1| (AE006574) putative  
biotin operon repressor [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAM79643.1"  
/db\_xref="GI:21904773"  
/translation="MKTSEKIYQLLSQTDQDFVSGEYLADOLISRTSVWKSIXLSLENO  
GIQIDSLRKHYNMVQGDILLPKTISQGLGMPVYTPPHSQSTQLDQAKQIEAHNSAPR  
LYLAPSOEAAKGRDLRQFFSASTGGIYMSYLPKNVPYADMPPTNMVAFVFPPEAISA  
LTGIDTEIKWNDIYLNHNVKAGILTAITSVETGLITDVIIGVGLNFVFTDPPPEAIA  
QKAGSLFTBKPTIRNDLIIDIWKLFLSIPVKDHVKVYKESLVLNKLQVTFIENSQOEK  
RAIADLTQDGHILVQFENGDLQTLRSGBISLSWEN"  
complement (8160.. .8444)  
/gene="SpyM3\_1037"  
complement (8160.. .8444)  
/gene="SpyM3\_1037"  
/note="best non-GAS blastp hit: gb|AAK74993.1| (AE007392)  
hypothetical protein [Streptococcus pneumoniae TIGR4], and  
gb|AAK99574.1| (AE008452) hypothetical protein

[Streptococcus pneumoniae R6]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM79644.1"  
/db\_xref="GI:21904774"  
/translation="MGKVLPLPTFVLTFFHSTVFFCYNKAMTIRQFLFMAFVCAFET  
YFNDLLSGNYLFALFWGLLFLDLRLRVHTINQLTKILKANSPPKKD"  
complement (8448.. .10118)  
/gene="dnaX"  
/note="synonym: SpyM3\_1038"  
complement (8448.. .10118)  
/gene="dnaX"  
/note="best blastp hit: gb|AAL97971.1| (AE010058) DNA  
polymerase III gamma/tau subunits [Streptococcus pyogenes  
MGAS8232]"  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAM79645.1"  
/db\_xref="GI:21904775"  
/translation="MYQALYRKYSQTFDEMVGQSVITTLKQAVESGKISHAYLPSFG  
PRGTGKTSAAKIFAKAMNCPNQVDGPCNQCDICRITNGSLEDVIEIDAASNNQVDE  
IRDIRKSTYAPSPFKAIRKQAIKREHLAWLKDGEIAYEVDALNLIARAEGMRDALS  
IPATILSRVQRFKAIKQAIKREHLAWLKDGEIAYEVDALNLIARAEGMRDALS  
ILDOALSIPDNQVTLIAEBITGTSISMALGDVYRVYSQOATCALALETIYDSGK  
SMSPATDILLTYLRDLLVYKAGDNQORSAYPDNLSLIDRIFQMTVTVSHLSEIK  
KQTHPRIYAEWMTIQLOKEQLSQANLSGELISIEITLKNELAQKQLQSLQSRPD  
SLAASDKTPKTSYRVDRTVLTIKMEETVRNSQOQRYLDALKNWNEILDNISQD  
RALLMGSEPLTANSYENALFAEAFNAQVMSNRNLMFGNIMSKAAGFSNLAAP  
RTDFQIRKEFAQKMSQKSVQEBEQEVALDIPEGDFLLDKINTIDD"  
complement (10118.. .10681)  
/gene="SpyM3\_1039"  
complement (10118.. .10681)  
/note="best non-GAS blastp hit: gb|AAK06294.1| (AE006448)\_12  
(AE006448) hypothetical protein [Lactococcus lactis subsp.  
lactis]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM79646.1"  
/db\_xref="GI:21904776"  
/translation="MTEAKGRKKYSNRPVAILKGMKSKKIEYOYLMIAQAKELFA  
NESNALANLSNALLNMTLNSVPAGFYLDGQELILGPFQGRVSCVHILKLGKVCQ  
ESASRRRTIINDVQKHANYISCDAAAMSEIVVPMVKEGHLIGVLDLSSLVADYDEV  
DQYLEAFVDFLEKTTTFNMGVKN"  
10760.. .11569  
/gene="SpyM3\_1040"  
10760.. .11569  
/note="best non-GAS blastp hit: gb|AAK98831.1| (AE008387)  
Query Match 100.0%; Score 21; DB 1; Length 51454;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACGTGATGGTTCCTTC 21  
|||||  
DB 18222 TGCATGTATGGTTCCTTC 18242  
RESULT 37  
BA000034\_07/c  
WPCOMMENT  
Sequence split into 19 fragments LOCUS BA000034 Accession BA000034  
Fragment Name Begin End  
BA000034\_00 1 110000  
BA000034\_01 100001 210000  
BA000034\_02 200001 310000  
BA000034\_03 300001 410000  
BA000034\_04 400001 510000  
BA000034\_05 500001 610000

BA000034\_06 600001 710000  
BA000034\_07 700001 810000  
BA000034\_08 800001 910000  
BA000034\_09 900001 1010000  
BA000034\_10 1000001 1110000  
BA000034\_11 1100001 1210000  
BA000034\_12 1200001 1310000  
BA000034\_13 1300001 1410000  
BA000034\_14 1400001 1510000  
BA000034\_15 1500001 1610000  
BA000034\_16 1600001 1710000  
BA000034\_17 1700001 1810000  
BA000034\_18 1800001 1894275  
Continuation (8 of 19) of BA000034 from base 700001 (BA000034 Streptococcus pyogenes SS1

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 107826 TGCATGTATGGGTATCTTCC 107806

RESULT 38  
BA000034\_08/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS BA000034 Accession BA000034  
Fragment Name Begin End  
BA000034\_00 1 110000  
BA000034\_01 100001 210000  
BA000034\_02 200001 310000  
BA000034\_03 300001 410000  
BA000034\_04 400001 510000  
BA000034\_05 500001 610000  
BA000034\_06 600001 710000  
BA000034\_07 700001 810000  
BA000034\_08 800001 910000  
BA000034\_09 900001 1010000  
BA000034\_10 1000001 1110000  
BA000034\_11 1100001 1210000  
BA000034\_12 1200001 1310000  
BA000034\_13 1300001 1410000  
BA000034\_14 1400001 1510000  
BA000034\_15 1500001 1610000  
BA000034\_16 1600001 1710000  
BA000034\_17 1700001 1810000  
BA000034\_18 1800001 1894275  
Continuation (9 of 19) of BA000034 from base 800001 (BA000034 Streptococcus pyogenes SS1

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 7826 TGCATGTATGGGTATCTTCC 7806

RESULT 39  
CP000003\_10  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000003 Accession CP000003  
Fragment Name Begin End  
CP000003\_00 1 110000  
CP000003\_01 100001 210000  
CP000003\_02 200001 310000  
CP000003\_03 300001 410000  
CP000003\_04 400001 510000  
CP000003\_05 500001 610000  
CP000003\_06 600001 710000  
CP000003\_07 700001 810000  
CP000003\_08 800001 910000

CP000003\_09 900001 1010000  
CP000003\_10 1000001 1110000  
CP000003\_11 1100001 1210000  
CP000003\_12 1200001 1310000  
CP000003\_13 1300001 1410000  
CP000003\_14 1400001 1510000  
CP000003\_15 1500001 1610000  
CP000003\_16 1600001 1710000  
CP000003\_17 1700001 1810000  
CP000003\_18 1800001 1899877  
Continuation (11 of 19) of CP000003 from base 1000001 (CP000003 Streptococcus pyogenes

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 92981 TGCATGTATGGGTATCTTCC 93001

RESULT 40  
CP000017\_10  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000017 Accession CP000017  
Fragment Name Begin End  
CP000017\_00 1 110000  
CP000017\_01 100001 210000  
CP000017\_02 200001 310000  
CP000017\_03 300001 410000  
CP000017\_04 400001 510000  
CP000017\_05 500001 610000  
CP000017\_06 600001 710000  
CP000017\_07 700001 810000  
CP000017\_08 800001 910000  
CP000017\_09 900001 1010000  
CP000017\_10 1000001 1110000  
CP000017\_11 1100001 1210000  
CP000017\_12 1200001 1310000  
CP000017\_13 1300001 1410000  
CP000017\_14 1400001 1510000  
CP000017\_15 1500001 1610000  
CP000017\_16 1600001 1710000  
CP000017\_17 1700001 1810000  
CP000017\_18 1800001 1838554  
Continuation (11 of 19) of CP000017 from base 1000001 (CP000017 Streptococcus pyogenes

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 96984 TGCATGTATGGGTATCTTCC 97004

RESULT 41  
CP000056\_11  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000056 Accession CP000056  
Fragment Name Begin End  
CP000056\_00 1 110000  
CP000056\_01 100001 210000  
CP000056\_02 200001 310000  
CP000056\_03 300001 410000  
CP000056\_04 400001 510000  
CP000056\_05 500001 610000  
CP000056\_06 600001 710000  
CP000056\_07 700001 810000  
CP000056\_08 800001 910000  
CP000056\_09 900001 1010000  
CP000056\_10 1000001 1110000  
CP000056\_11 1100001 1210000

CP000056_12	1200001	1310000
CP000056_13	1300001	1410000
CP000056_14	1400001	1510000
CP000056_15	1500001	1610000
CP000056_16	1600001	1710000
CP000056_17	1700001	1810000
CP000056_18	1800001	1897573
Continuation (12 of 19) of CP000056 from base 1100001 (CP000056 Streptococcus pyogenes M		
Query Match	100.0%;	Score 21; DB 1; Length 110000;
Best Local Similarity	100.0%;	Pred. No. 5.2;
Matches 21; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGCATGTATGGGTATCTTCC 21
Db	20289	TGCATGTATGGGTATCTTCC 20309
RESULT 42		
SAG766847/C		
LOCUS	SAG766847	167050 bp DNA linear BCT 16-APR-2005
DEFINITION	Streptococcus agalactiae NEM316 complete genome, segment 5.	
ACCESSION	AL766847	AL732656
VERSION	AL766847.1	GI:23095200
KEYWORDS		
SOURCE	Streptococcus agalactiae NEM316	
ORGANISM	Streptococcus agalactiae NEM316	
REFERENCE	1	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
AUTHORS	Glaser, P., Rueniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Meadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P. and Kunst, F.	
TITLE	Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease	
JOURNAL	Mol. Microbiol. 45 (6), 1499-1513 (2002)	
PUBMED	12354221	
REFERENCE	2	Glaser, P., Rueniok, C. and Frangeul, L.
AUTHORS	Direct Submission	
TITLE	Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86	
JOURNAL	Location/Qualifiers	
FEATURES	1. .167050	
source	/organism="Streptococcus agalactiae NEM316"	
	/mol_type="genomic DNA"	
	/strain="NEM316"	
	/serovar="serovar III"	
	/db_xref="taxon:211110"	
gene	join(288. .293,303. .1412)	
RBS	288. .293	
	/genes="gbs0767"	
CDS	/genes="gbs0767"	
	303. .1412	
	/genes="gbs0767"	
	/note="Similar to riboflavin specific deaminase (diaminoxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil reductase)"	
	/codon_start=1	
	/transl_table=11	
	/product="Unknown"	
	/protein_id="CAD46411.1"	
	/db_xref="GI:23095201"	
	/db_xref="GOA:Q8E660"	
	/db_xref="InterPro:IPR002125"	
	/db_xref="InterPro:IPR002734"	
	/db_xref="InterPro:IPR004794"	
	/db_xref="InterPro:IPR011549"	
	/db_xref="UniProt/TREMBL:Q8E660"	
	/translation="MMKEDYNALALKEAEGKMGFVAPNPLVGAVIVKDDRIISKGYH	

KRFGLHAEROAIKNADEDISGSTLYVTLEPCCHVGKQPCTEALIKSGIKKVVVGSLL	gene
DPNPLVSGKGTALLRKECLNVEVGIILRECDALNERFIHMTYKQPFVYLKYAMTLDG	RBS
KIATKIGDSKWIISNEHSQSVQKLRQKCSAMVGINTVLADNPRLTCTIPKGEALVRI	gene
VCDSQLKIPLDSILYKSAKTIPTWTATCDNLAAQQOITLKEMCGRLIKVPRKQKLDLK	CDS
VLMITLIGQEGIDSLIIEGSSLSLHFSALRAGIVNRLIVFIAPKIIIGLKAKTAISGEG	
DWLNQDFRVKDILSRMDSVVIEGKVEHYVYRN	
1381. .1387	
/gene="gbs0768"	
1381. .1387	
/gene="gbs0768"	
1393. .2043	
/gene="gbs0768"	
1393. .2043	
/gene="gbs0768"	
/note="Similar to riboflavin synthase alpha chain"	
/codon_start=1	
/transl_table=11	
/product="Unknown"	
/protein_id="CAD46412.1"	
/db_xref="GI:23095202"	
/db_xref="InterPro:IPR001783"	
/db_xref="UniProt/TREMBL:Q8B659"	
/translation="MFTGIIEMGVSRIRNGIKSQQLSIDAPKLVLPLLRKGDSSVAVN	
GVCLTVLDKSETAFIADVPESMMRTSLAALRLHSLKVNLEALRSDSLGCHFVIGHV	
DGVGKIEIKOKDIAVRFSIDAPPSIMSYIIKGSVALDGLSLTVVSTFHSFEVSVI	
PHTMAQTNLSLKKVGDLLNIEVDVLGKYAEKFLAPTNRNHTSSVMDMSFLSENGY"	
2046. .3254	
/gene="gbs0769"	
2046. .2051	
/gene="gbs0769"	
2061. .3254	
/gene="gbs0769"	
/note="Similar to GTP cyclohydriase II/ 3,4-dihydroxy-2-butanone-4-phosphate synthase"	
/codon_start=1	
/transl_table=11	
/product="Unknown"	
/protein_id="CAD46413.1"	
/db_xref="GI:23095203"	
/db_xref="GOA:Q8B658"	
/db_xref="InterPro:IPR000422"	
/db_xref="InterPro:IPR000926"	
/db_xref="UniProt/TREMBL:Q8B658"	
/translation="MAFSPIKKLQDIKSGKVMVMDDENRENGDLICAAEMVTKES	
INFMAKFGKGLICLPLSNYYAEKLELAQMASHTNDNHTAFTTISIDHLSSTGICSAED	
RALTAMVANDSSKADPRRPHLPFLLAKEGGVLARNGHTAATVDLCRLAGLKCCGL	
CCEINAEDEGSMWRKDELLAFQKHDLAITIKQDYRRQEGGVVRIEIQLPQFG	
HFTAYGSEVVANKHEVALVRGDISSGEDVLCRLHSECLTGDVPHSLRCDCEQLANA	
LQQTBAEGRGVLLYMRQGRGIGLINKLKYALOEGLDTLEANLALFEGEDRYGV	
SAQLLKDLGINSINLLTNPNPKIQQLEAEGICVKNRVPLQVAATAAYDLNLYLTKKEKM	
GHLDD"	
3255. .3739	
/gene="gbs0770"	
3255. .3260	
/gene="gbs0770"	
3269. .3739	
/gene="gbs0770"	
/note="Similar to riboflavin synthase complex beta chain"	
/codon_start=1	
/transl_table=11	
/product="Unknown"	
/protein_id="CAD46414.1"	
/db_xref="GI:23095204"	
/db_xref="GOA:Q8B657"	
/db_xref="InterPro:IPR002180"	
/db_xref="UniProt/Swiss-Prot:Q8B657"	
/translation="WTIIIEGQLVANEMKIGIVWSRFNELITSKLLSGAVDGLLRHGVS	
BEDIDIVWPGAFIIPYNARKMVKDYDAIICLGWIKGSTDHYVVCNEVTKIGH	
LNSQSDIPIHFGVLTLDNIEQALIERAGTKAGNKGYDCALSAIEMVNLDKLER"	
complement (3814. .5315)	
/gene="lyss"	
complement (3814. .5304)	
/gene="lyss"	

```
/codon_start=1
/transl_table=11
/product="lysyl-tRNA synthetase"
/protein_id="CAD46415.1"
/db_xref="GI:23095205"
/db_xref="GOA:Q8E656"
/db_xref="InterPro:IPR002313"
/db_xref="InterPro:IPR004364"
/db_xref="InterPro:IPR004365"
/db_xref="InterPro:IPR006195"
/db_xref="InterPro:IPR008994"
/db_xref="UniProt/Swiss-Prot:Q8E656"
/translation="MNSQIHELNDQIVREKMAALTEQIDPFGKRFERTATSQOL
NEKYADKSDIEDHDIETETIAGRLMTKRGKGVFAHIQDREGQIQIYVRKDSVGE
NVEIFPKADLGLVEGQVMTDMGELSIKATHITLSKALRPLPEKFGHTDIETI
YRKRHLDSNRDSFQVTRSKIISEIRFMSDNGFLEVETPVLNHEAGASARFPI
THNAQIDIMVLRIATLHLKRLIVGGMERVVEIGRIFRNEGMDATHNPEFTSEIAYQ
AVADYQIDMLTGTIQTHTKTVKGDPINQGTIEIKINEPKFRHMDVDAKVEITGID
FKEMTLEBAQAQKQVPLEKHFTTVGHIINAFEEFVEDTLIQTFVFGHPVEVS
FLAKNDTPRFTRDELFTMTKEYANAFTELNDPIDQLSRPEAQASAKELGDEATG
VDYVYEALEYGMPPGGIGIGIDRLCMLLTDTTIRDVLLFFPMK"
complement(5310..5315)
/gene="lyeS"
5467..6411
/gene="gbs0772"
5467..5472
/gene="gbs0772"
5479..6381
/gene="gbs0772"
/notes="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAD46416.1"
/db_xref="GI:23095206"
/db_xref="GOA:Q8E655"
/db_xref="InterPro:IPR005833"
/db_xref="InterPro:IPR005834"
/db_xref="InterPro:IPR006439"
/db_xref="UniProt/TREMBL:Q8E655"
/translation="MTSIVFDVDDTIYDQAPYRIAVEKCFDPFDMASINQAIYRFR
HYSIDIGPRVMAGEVTEYFRFWCKETLLLEFGVREIDEATGIYFQBIYHELENIITM
LDSMRWTLDFLKSQVPMGIITNGPTEHQLKVKKGLGYDYDPKRVIVSQATGFKQPK
EKIFNLAAQFQMNSTTLVYGDSDVNDLWGAFNGHSMWFNHRGRSLKPGIKPVY
DVALDNFEQLFGAVKVLFDLPDNKFIIDINDKSNPULMGLNGLNGLMMAERLLESNMS
VDKVVILLRLTAQKQKVLRMKYAR"
6390..6411
/gene="gbs0772"
complement(6417..7069)
/gene="gbs0773"
complement(6417..7058)
/gene="gbs0773"
/notes="Similar to putative phosphoglycerate mutase"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAD46417.1"
/db_xref="GI:23095207"
/db_xref="GOA:Q8E654"
/db_xref="InterPro:IPR001345"
/db_xref="UniProt/TREMBL:Q8E654"
/translation="MKFYLRHGKQWNLEGRFGQANGDSPLLEEAIEELEELGOVLS
SIHPDAVSSDLGKARDVINLNDASCPKEIHYTPQLRWALGTGLGCKIATQWAIY
PQWMTYQNPLOQFDMFGAESLYQTTHREVSFLSKNSKLYDKVLIVGHGANLTAS
IRSLYGQYGLSHYKDKLDNASLTIIETHDFKDFNCLTWNDKSYLROEVKWTM"
complement(7064..7069)
/gene="gbs0773"
complement(7080..7553)
/gene="gbs0774"
complement(7080..7553)
/gene="gbs0774"
/notes="Similar to transcriptional regulator"
/codon_start=1
```

```
/transl_table=11
Query Match 100.0%; Score 21; DB 1; Length 167050;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 80038 TGCATGTATGGGTATCTTCC 80018

RESULT 43
CO655069/c 349980 bp DNA linear PAT 02-FEB-2004
LOCUS
DEFINITION
Sequence 12026 from Patent WO0234771.
ACCESSION
CO655069
VERSION
CO655069.1 GI:41687946
KEYWORDS
Streptococcus agalactiae
SOURCE
Streptococcus agalactiae
ORGANISM
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Telford,J., Maignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE
Nucleic acids and proteins from streptococcus groups a & b
JOURNAL
Patent: WO 0234771-A 12026 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1..349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/notes="seq 10967 to big: splitted in
seq 10.967, from 0.000.001 to 0.349.980
seq 12.025, from 0.300.001 to 0.649.980
seq 12.026, from 0.600.001 to 0.949.980
seq 12.027, from 0.900.001 to 1.249.980
seq 12.028, from 1.200.001 to 1.549.980
seq 12.029, from 1.500.001 to 1.849.980
seq 12.030, from 1.800.001 to 2.149.980
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 191815 TGCATGTATGGGTATCTTCC 191795

RESULT 44
AX954529/c 349980 bp DNA linear PAT 08-JAN-2004
LOCUS
DEFINITION
Sequence 1375 from Patent WO03093306.
ACCESSION
AX954529
VERSION
AX954529.1 GI:40783902
KEYWORDS
Streptococcus pyogenes
SOURCE
Streptococcus pyogenes
ORGANISM
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Telford,J., Maignani,V., margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE
Nucleic acids and proteins from streptococcus groups a b
JOURNAL
Patent: WO 03093306-A 1375 13-NOV-2003;
Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1..349980
/organism="Streptococcus pyogenes"
```

```

/mol_type="unassigned DNA"
/db_xref="taxon:1314"
/notes="seq 1373 to long, original length: 2.160.266
replaced by
seq 1373: from 0.000.001 to 0.349.980
seq 1374: from 0.300.001 to 0.649.980
seq 1375: from 0.600.001 to 0.949.980
seq 1376: from 0.900.001 to 1.249.980
seq 1377: from 1.200.001 to 1.549.980
seq 1378: from 1.500.001 to 1.849.980
seq 1379: from 1.800.001 to 2.149.980
seq 1380: from 2.100.001 to 2.160.266"

ORIGIN

Query Match          100.0%; Score 21; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 4.3;          0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
|||||
Db 220602 TGCATGTATGGGTTATCTTCC 220582

RESULT 45
AX569143/c
LOCUS AX569143 402 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 2351 from Patent WO02077021.
ACCESSION AX569143
VERSION AX569143.1 GI:26002638
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
1 Maignani, V., Tettelin, H. and Fraser, C.
Streptococcus pneumoniae proteins and nucleic acids
Patent: WO 02077021-A 2351 03-OCT-2002;
Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
LOCATION/Qualifiers
1. .402
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"

ORIGIN

Query Match          92.4%; Score 19.4; DB 6; Length 402;
Best Local Similarity 95.2%; Pred. No. 87;          0; Indels 1; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
|||||
Db 377 TGCATGTATGGGTTATCTTCC 357

RESULT 46
AR480287/c
LOCUS AR480287 1140 bp DNA linear PAT 14-MAY-2004
DEFINITION Sequence 1191 from patent US 6699703.
ACCESSION AR480287
VERSION AR480287.1 GI:47239249
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
1 (bases 1 to 1140)
Doucette-Stamm, L., Bush, D., Zeng, Q., Opperman, T. and
Houseweart, C.E.
Nucleic acid and amino acid sequences relating to Streptococcus
pneumoniae for diagnostics and therapeutics
Patent: US 6699703-A 1191 02-MAR-2004;
Genome Therapeutics Corporation; Waltham, MA
LOCATION/Qualifiers

REFERENCE
1 (bases 1 to 1140)
Doucette-Stamm, L., Bush, D., Zeng, Q., Opperman, T. and
Houseweart, C.E.
Cloning, sequencing and expression in Escherichia coli of the ptsi
gene encoding enzyme I of the phosphoenolpyruvate:sugar
phosphotransferase transport system from Streptococcus salivarius
Gene 121 (1), 71-78 (1992)
1427100
PUBMED
COMMENT Original source text: Streptococcus salivarius (library: ATCC
25975) DNA.

FEATURES
source
1. .2259
Location/Qualifiers
/organism="Streptococcus salivarius"
/mol_type="genomic DNA"
/db_xref="taxon:1304"

ORIGIN

Query Match          92.4%; Score 19.4; DB 6; Length 1287;
Best Local Similarity 95.2%; Pred. No. 72;          0; Indels 1; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
|||||
Db 431 TGCATGTATGGGTTATCTTCC 411

RESULT 48
STRIPBPA/c
LOCUS STRIPBPA 2259 bp DNA linear BCT 26-APR-1993
DEFINITION Streptococcus salivarius phosphoenolpyruvate:sugar
phosphotransferase system enzyme I (ptsi) gene, complete cds.
ACCESSION M81756.1 GI:153614
VERSION M81756
KEYWORDS cytoplasmic protein; phosphoenolpyruvate:sugar phosphotransferase
system enzyme I; protein kinase; ptsi gene; sugar transport
protein.
SOURCE Streptococcus salivarius
ORGANISM Streptococcus salivarius
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 2259)
Gagnon, G., Vadeboncoeur, C., Levesque, R.C. and Frenette, M.
Cloning, sequencing and expression in Escherichia coli of the ptsi
gene encoding enzyme I of the phosphoenolpyruvate:sugar
phosphotransferase transport system from Streptococcus salivarius
Gene 121 (1), 71-78 (1992)
1427100
PUBMED
COMMENT Original source text: Streptococcus salivarius (library: ATCC
25975) DNA.

FEATURES
source
1. .2259
Location/Qualifiers
/organism="Streptococcus salivarius"
/mol_type="genomic DNA"
/db_xref="taxon:1304"

ORIGIN

```

gene	/tissue_lib="ATCC 25975"									
-35_signal	22..2023									
TATA_signal	/gene="ptcI"									
CDS	22..27									
	/gene="ptcI"									
	47..52									
	/gene="ptcI"									
	79..1812									
	/gene="ptcI"									
	/codon_start=1									
	/transl_table=11									
	/product="phosphoenolpyruvate:sugar phosphotransferase									
	system enzyme 1"									
	/protein_id="AAA26873.1"									
	/db_xref="GI:153615"									
	/translation="MTMLKGIASDGVAVAKAYLLVQPDLSFETVTVEDTSABEARL									
	DALAAASQDELSVIREKAVESLGEAAAVFDALHMLVADPEMTGQIKETIRAKQVNAE									
	AALTEVDMFIALFEGMEDNPYMQEAAADIRDVTKRVLANLIGKLPNPATINEESIV									
	VAHDLTPSDTAQLNKYKAFVNIIGRTSHSAIMARTLEIAAVLTGNNITELVQGD									
	ILAVSGITGEVINPTEQIAEFKAAGEAYAKQAEWALLKOAQTVTAGKHPFLAAN									
	IGTPKQVGVNDNGAEGVLRTEFLYDMSQDPPTDDQYEAQKAVLEGMGKPVVVR									
	TMDIGGDKELPYFDLPKEMNPFILGYRALRISISETGNQMFRTQLRALLRASVHGKURI									
	MFPWALLTEFRITAKGILBEKAKLVAEGVAVADDIEVGIMIEIPAAAMLADQFAKEV									
	DFPSIGTNDLIQVTMAADRMEQSVLYQPNPSILRLINNVIAAHABGKWAGMCGE									
	MAGDOTAVPLLVGMLDFEFSMSATSVLRTSLMKKLDTAQMEYANRALTECSTMEEV									
	LELSKEYVNDV									
terminator	1817..1847									
	/gene="ptcI"									
terminator	/note="putative"									
	1996..2023									
	/gene="ptcI"									
	/note="putative"									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										
b	455 TGCATGTATGGTGTCTTCC 435									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Q	1 TGCATGTATGGTGTATCTTCC 21									
Y										





SKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBEILMKDPNPKDKED  
IYNEVGKGYIVKDGKYYVLYKDAADNVRUTKEEINROKQHSQHREGSTPNDGAV  
ALARSQRYITDGGYIFNADSIITEDGDAIVPHGHDHYHI PKNELSASELAAAEFL  
SGRGNLSRKYTRQMSDTRTNWVPSVNSPGTTNTNNSNTNSQASNDIDSL  
LKQLYPLSRHVEGSDGLVDFPAQITSRTARGVAPHGDHYHFIPIYSQMSLEERIA  
RIIPLRYRNSWVPDRPEQSPQPTPEPSPQAPAPNLKIDNSLSVLQVRKVGEG  
YVFEKIGSYFAKOLPSLTVKNLESKLSQBSVSHTLTAKKENVAPRQDFYDKAY  
NLTAEAKALFENKGSNSFOALDKLLERLNDESTNKEKLVDLLAPLPIHTPERLG  
KPNQSYTEDEVRIAQLADKYTTSGYIFDEHDIIISDEGDVYTPHMGHSHWIKDS  
LSDEKVAAGQYKTEGILPPSPADVKANFTGDSAAIYTRNVKGEKRIPLVRLPVMY  
EHTVEKNGNLIIPKHDYHNIIPAFMDHDHYKAPNGYTLIEDLFIKYVVEHRRP  
HNSDNGWNASVHLKGDHSEDPKNKFKADEEPEVETPAEPVQVETEKVEAQLEKA  
EVLALYTDSSKANATETLAGLRNNLTQLQMDNNSIMAEAKLLALLKGSNSPSSVK  
EKIN"

gene  
complement (8049. .9782)  
/gene="SP1176"  
CDS  
complement (8049. .9782)  
/gene="SP1176"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate-protein phosphotransferase"  
/protein\_id="AAK75285.1"  
/db\_xref="GI:14972656"

/organism="MTEMLKGIASDGVAVAKAYLIVQPLDSPETITVEDTNAEERL  
/translations="MTEMLKGIASDGVAVAKAYLIVQPLDSPETITVEDTNAEERL  
DAALQASDELVSIREKAVCTLGEAAQVDAHLMVLADPEMISQIKETIRAKKNVAE  
AGLKEVTDMITFPEGMEDPNYQERAADIRDVTKVLANLGLKLPNPASINEEVIV  
IAHDLTSDAQDKNFVKAPVTNIGGRTSHSAIMARTLEIAAVLGTNNITEIVKGD  
ILAVNGTSTVINTPDEQAAEFKAAGEAYAKQAEWALLKDAQTVDAGKHFEELAN  
IGTPKQVEGVNNGAEAVGYLRTFELYMDSQDFTEDEQYAYKAVLGNMGKPVVVR  
TMDIGGKELPYFDMPEHMPFLGFRALRISISETGDAMFTQIRALLRASVHGQLRI  
MFPWALLKEFRAAKAVDEEKANLLAEGVAVADNIQVIMIEIPAAAMLADQFAKEV  
DFPSIGNDLIQYTMADRMNEQSVLYQYNPISILRLINNVIKAAHAEKGWAGMCE  
MAGDQQAPLLVGLGGLDEFMSATSVLRTSLMKGLDTAKMBEYANRALTECSTMBEV  
LELQKEYVNF"

gene  
complement (9788. .10051)  
/gene="SP1177"  
CDS  
complement (9788. .10051)  
/gene="SP1177"  
/codon\_start=1  
/transl\_table=11  
/product="phosphocarrier protein Hpr"  
/protein\_id="AAK75286.1"  
/db\_xref="GI:14972657"  
/translation="MASKDFHVAETGIIHARPATLLVQTASKPASDITLEYKGSVNL  
KSLNGVMSLGVGGQADVTTISAEGADADDAIAAISETMEKEGLA"

ORIGIN

Query Match 92.4%; Score 19.4; DB 1; Length 10256;  
Best Local Similarity 95.2%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTCCTTC 21  
|||||  
Db 9406 TGCATGTATGGGTTCCTTC 9426

RESULT 53  
AE008479  
LOCUS  
DEFINITION Streptococcus pneumoniae R6 section 95 of 184 of the complete genome.  
AE008479 10320 bp DNA linear BCT 13-SEP-2001

ACCESSION AE008479 AE007317  
VERSION AE008479.1 GI:15458677  
KEYWORDS  
SOURCE Streptococcus pneumoniae R6  
ORGANISM Streptococcus pneumoniae R6  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 10320)  
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,  
DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,  
Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,  
LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,  
Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rocky,P.,  
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,  
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.  
and Glass,J.I.  
Genome of the bacterium Streptococcus pneumoniae strain R6  
J. Bacteriol. 183 (19), 5709-5717 (2001)  
11544234

2 (bases 1 to 10320)

Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,  
DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,  
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,  
LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,  
McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,  
Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rocky,P.,  
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,  
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.  
and Glass,J.I.  
Direct Submission

Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and  
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA  
Company, Location/Qualifiers

FEATURES

source

1. .10320  
/organism="Streptococcus pneumoniae R6"  
/mol\_type="genomic DNA"  
/strain="R6"  
/db\_xref="taxon:171101"

gene

complement (91. .738)  
/gene="spr1056"  
CDS  
complement (91. .738)  
/gene="spr1056"  
/codon\_start=1  
/transl\_table=11  
/label="spr1056"

/product="Hypothetical protein"  
/protein\_id="AAK99860.1"  
/db\_xref="GI:15458678"

/translations="MTTKANITIEEYIEMSEVDFNEAVNVEFTSDTCQLANSTYQSL  
FKFDFKNFSGDLIFTWKPSLVKEGDIYGRDRSQVDNLRAVIGNIFNLTKRYSLN  
MNRNGCMGDFPHDFDIYLDHVAKYAEQKNVKEYPKLRALLHOENALYPRFESN  
FDDFLKYNLTKTIWQSKETPPSFDFNMFKNISEKIIFERGSKMLNDLSNKKK"  
complement (752. .1465)  
/gene="spr1057"

gene

CDS  
complement (752. .1465)  
/gene="spr1057"  
/codon\_start=1  
/transl\_table=11  
/label="spr1057"

/product="Conserved hypothetical protein"  
/protein\_id="AAK99861.1"  
/db\_xref="GI:15458679"

/translations="MFYKFLFDLHDTLLDPAEDVALTOLLKEGVADIQAYKDYV  
VPMKALWDLLEKKISKQELVNTFRSLRSHFGQEGDGLAORYQFYLAQQGTL  
GAHDLSDLIERYDYLYATNGITAIQTGLAOSGLVYFNQVFISSQLQKQDALL  
YEKIQOIAFGSKETLMIGDSLTDITQGGNNAGIDTIWYNPHPLENHTQAPTYEVY  
SYDLDLDDKNILEKITP"  
complement (1923. .2081)  
/gene="spr1058"

gene

CDS  
complement (1923. .2081)  
/gene="spr1058"  
/codon\_start=1  
/transl\_table=11  
/label="spr1058"

/product="Hypothetical protein"  
/protein\_id="AAK99862.1"  
/db\_xref="GI:15458680"

/translations="MNLKWTDFSDKSIINRVPIVEGAYNAAALQAGAELSVIQTO  
LAELINK"

gene

complement (2151. .2306)  
/gene="spr1059"  
CDS  
complement (2151. .2306)  
/gene="spr1059"  
/codon\_start=1

```
/transl_table=11
/label=spr1059
/product="Hypothetical protein"
/protein_id="AAK99863.1"
/db_xref="GI:15458681"
/translation="WMESIGIGLIVIVSHSKHIAEGVVELISKVAKDVPITYVGGTEGG
GIGT9FD"
repeat_region
complement (2542..2613)
/rpt_family="RUP element"
/rpt_type="dispersed"
/rpt_unit="rup_a"
complement (2725..5292)
/gene="pnpA"
/notes="synonym: spr1060"
CDS
/gene="pnpA"
complement (2725..5292)
/gene="pnpA"
/codon_start=1
/transl_table=11
/label=spr1060
/product="Histidine Motif-Containing protein"
/db_xref="GI:15458682"
/translation="MNQIYLRKEERMKINKYLAGSVATLVLSVCAYELGLHQAOTVK
ENNRVSYIDGKATQKTENLTPTDEVSKREGINAQIVIKITDQGYVTSBGHGHYNYG
KYPYDAIIEELLMKDPNYQLKDEDIISIKGGYIVKDGKYYVLKDAAHADNVRTK
EENRQKQSHSQRHREGGTRNDGAVALARSGQRYTTDDGYIFNASDI EDTGDATVP
LGDHGHYIIPKNELASELAAAKAFLSGRGNLSNRTYRRQNSDNTSRTNVSVNPG
TTNTNSNNTNSQASNDIDSLKQYIKLPLSORHVESDGLIFDPAQITSRANG
VAVPDGHVHFIPYSQLSPLEEKLARIIPLYRSNHWPPDSRPEQSTPEPSPKL
QAPNPQAPNPIDSKLVEAKVRKVDGYFEENGVPRIYIPAKOLSAFTAAGIDSKL
AKQESLHKLGAKTLDTPSSDRFYNKAYDLLARIHQDLDNKGROVDFEALDNLIER
LKDVSSDKVLKVDLILAFIAPLRHPERLKGPNQAIIYTDEIQVAKLACKYTTEDGYI
FPRLDITSEGDATVYTPMTHSHWIKDLSI SEARAAQAAYAKEKGLTPPSDHDQSG
NTEAKGAIINRWKAAKVPLODRMEYNLQYTVVEVNGSLIIPHDHNNHINKFNPDE
GLYEAPKSYLDELATVYVYVEHNERPHSDNGFNGASDHVQRNKNGQADNTQTEKP
NEEKPQTEKEDPREEPKQSEKPSKPTEPESEPESEPESEPESEPESEPESEPESEPE
KLREADLLGKIQNPITIKSNAKETLTGLKNLLFGQDNNTIMAEAKLLALLKESK"
complement (5417..7903)
/gene="pntA"
/notes="synonym: spr1061"
CDS
complement (5417..7903)
/gene="pntA"
/codon_start=1
/transl_table=11
/label=spr1061
/product="Pneumococcal histidine triad protein A
precursor"
/protein_id="AAK99865.1"
/db_xref="GI:15458683"
/translation="MQLEISNRKVSVMKINKYLVGSAAALILSVCSVELGYQARTV
KENNRVSYIDGKATQKTENLTPTDEVSKREGINAQIVIKITDQGYVTSBGHGHYNY
KGPYDAIIFSELLMKDPNYQLKDEDIIVNEKGGYIVKDGKYYVLKDAAHADNVRT
KEEINRQKQSHSQRHREGGTRNDGAVALARSGQRYTTDDGYIFNASDI EDTGDATVP
PGDGHYIIPKNELASELAAAEFLSGRGNLSNRTYRRQNSDNTSRTNVSVNPG
GTTNTNSNNTNSQASNDIDSLKQYIKLPLSORHVESDGLVFDPAQITSTAR
GVAPHGDHGHYIIPYSOMSELERIARIIPLYRSNHWPPDSRPEQSPQPTPEPSGQ
POAPNLKDSNLSVQLVRKVGEGYFEKGISYVFAKOLPSRPNKLSKESKPK
ESVSHLTATKANVAPRDQFYDKAYNLLTEAHKALFENKGRNSDFQADKLLERLND
ESTNKEKLVDDLAFIAPLRHPERLKGPNQIYEYTEDVRIQALADKYTTSDGYIPDE
HDIISDEGATVYTPMTHSHWIKDLSI SEKVAQAQYAKEKGLTPPSDADVKANPT
GSAAIYNRVKGEKIPLVRLPYMVEHTVEVNGNLIIPKDHVHNIKFAWFDDBHTY
KAPNGTYLEDLATIKYVVEHDERPHSDNGWGNASEHVLGKHDSHEDKNKPKADEE
PVETPAPEVQVETKEVBAQKEARVLLAKVTDSLSKANATETIAGURNLITLIQM
DNNISMAEAKLALLKGNPSVSKSKIN"
complement (8122..9855)
/gene="ptsI"
/notes="synonym: spr1062"
CDS
complement (8122..9855)
/gene="ptsI"
/EC_number="2.7.3.9"
/codon_start=1
/transl_table=11
/label=spr1062
```

```
/product="PTS enzyme I"
/protein_id="AAK99866.1"
/db_xref="GI:15458684"
/translation="WTEMLKGIAASDGVAVAKAYLLVOPDLSFETITVEDTNAEBARL
DAALQASDELSVIREKAVGTLGEEAAQFODHLMVLADPFIKQITIRAKKNVAE
AGLKEVTDMFITIFEGMEDNPNQBRADIRDVTKRVLANLLGKLPNPINNEEVIV
IAHDUTPSDITQDLNDFYKAFVTNIGRTHSAIIMARTLEIAAVGTGNNITIEIVKDG
ILAVNGITGEVILNPTDQAAEFKAAGBAYAKQAEWALLKDAQVTADGKHFEALAN
IGTPKDGVEGNNGAEVGLYRTEFLYNDSDPTFEDQYEAAYKAVLEGMNCKPVVVR
TMDIGDKELPYFDMPHENPFLGRALRISISSETGDAMFRTQIBALLRASVHGOLRI
MPFWALLKEPFAAKAVFDEBKANLIASGVADNIIQVGMIEIPAAAMLAADQFAKEV
DFFSIGTNDLIQYTMAADRMEQSYLYQPNYPSILRLINNVIKAAHAEGRKWAGMCE
MAGDQAVPLLVGMGLDSEFSMSATSVLRSMLMKKLDTAQMEYANRALTEGSTMBEEV
LELOKEYVNF"
complement (9861..10124)
/gene="ptsH"
/notes="synonym: spr1063"
CDS
complement (9861..10124)
/gene="ptsH"
/codon_start=1
/transl_table=11
/label=spr1063
/product="Histidine-containing phosphocarrier protein of
the PTS"
/protein_id="AAK99867.1"
/db_xref="GI:15458685"
/translation="WASKDFHVVAETGTHARPATLLVOTASKFASDITILEYKKGKSVNL
KSIMGVMSLGVGGQADVTISAGADADDAAIAISETMEKGLA"
ORIGIN
Query Match 92.4%; Score 19.4; DB 1; Length 10320;
Best Local Similarity 95.2%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCACGTATGGTTCCTTC 21
|||||
Db 9479 TGCATGTATGGTTCCTTC 9499
RESULT 54
SPNEU1915/c
LOCUS
DEFINITION Streptococcus pneumoniae clone GS4, *** SEQUENCING IN PROGRESS ***.
ACCESSION AL449937
VERSION AL449937.1 GI:11545162
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE
1 Streptococcus.
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 13F clinical isolate
JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
PUBMED 11442348
REFERENCE 2 (bases 1 to 20035)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
Direct Submission
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
source
1..20035
```

```

/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"

ORIGIN
Query Match      92.4%; Score 19.4; DB 14; Length 20035;
Best Local Similarity 95.2%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 15401 TGCATGTATGGGTGTCTTCC 15381

RESULT 55
AX5711763
LOCUS      AX5711763      349980 bp      DNA      linear      PAT 29-NOV-2002
DEFINITION      Sequence 4982 from Patent WO02077021.
ACCESSION      AX5711763
VERSION      AX5711763.1 GI:26003955
KEYWORDS
SOURCE      Streptococcus pneumoniae
ORGANISM      Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1
REFERENCE
AUTHORS      Masignani,V., Tettelin,H. and Fraser,C.
TITLE      Streptococcus pneumoniae proteins and nucleic acids
JOURNAL      Patent: WO 02077021-A 4982 03-OCT-2002;
Chiron Spa (it); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1 .349980
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"
/notes="seq 4979 too long: 2.162.598 bases
replaced by following seq:
seq 4979: from 0.000.001 to 0.349.980
seq 4980: from 0.300.001 to 0.649.980
seq 4981: from 0.600.001 to 0.949.980
seq 4982: from 0.900.001 to 1.249.980
seq 4983: from 1.200.001 to 1.549.980
seq 4984: from 1.500.001 to 1.849.980
seq 4985: from 1.800.001 to 2.149.980
seq 4986: from 2.100.001 to 2.162.598"

ORIGIN
Query Match      92.4%; Score 19.4; DB 6; Length 349980;
Best Local Similarity 95.2%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 108618 TGCATGTATGGGTGTCTTCC 108638

RESULT 56
AP001046
LOCUS      AP001046      158063 bp      DNA      linear      PRI 14-JAN-2000
DEFINITION      Homo sapiens genomic DNA, chromosome 21, clone:KB43F12, MX1-D21S171
region, complete sequence.
ACCESSION      AP001046
VERSION      AP001046.1 GI:6693596
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
1 (bases 1 to 158063)
Shimizu,N., Kudoh,J. and Shibuya,K.

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/clone="KB43F12"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"

ORIGIN
Query Match      87.6%; Score 18.4; DB 8; Length 158063;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 20
    |||||
Db 90265 TGCATGTATGAGTTATCTTC 90284

RESULT 57
AC144901/c
LOCUS      AC144901/c      178293 bp      DNA      linear      HTG 22-JUL-2003
DEFINITION      Sus scrofa clone RP44-138L19, WORKING DRAFT SEQUENCE.
ACCESSION      AC144901
VERSION      AC144901.3 GI:331112720
KEYWORDS      HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Sus scrofa (pig)
ORGANISM      Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
1 (bases 1 to 178293)
REFERENCE
AUTHORS      Antonellis,A., Aylele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnov,M.E.,
Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantripp,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 178293)
Green,E.D.
Direct Submission
Submitted (24-MAY-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 178293)
Green,E.D.
Direct Submission
Submitted (22-JUL-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jul 22, 2003 this sequence version replaced gi:32306561.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information

```

Center project name: ecc  
Center clone name: 138L19

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig, has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 178287 bases at least Q40  
Consensus quality: 178293 bases at least Q30  
Consensus quality: 178293 bases at least Q20  
Insert size: 167000; agarose-fp  
Insert size: 178293; sum-of-contigs  
Quality coverage: 11.38x in Q20 bases; agarose-fp  
Quality coverage: 10.66x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 178293: contig of 178293 bp in length.

## FEATURES

## source

1..178293  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="RP44-138L19"  
/clone\_lib="RP44"

## misc\_feature

1..178293  
/note="assembly\_fragment  
clone\_end:Sp6  
vector\_side:left  
clone\_end:T7  
vector\_side:right"

## misc\_feature

1..83454  
/note="Clone overlaps with GenBank Accession Number  
AC145413 clone RP44-386P23 (center project name ecb)"

## ORIGIN

Query Match 87.6%; Score 18.4; DB 14; Length 178293;  
Best Local Similarity 95.0%; Pred.No. 1e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 1;

QY 1 TGCATGTCATGGGTTATCTTC 20

Db 156133 TGCATGTCATGGGTTACCTTC 156114

## RESULT 58

AC018734/c

LOCUS AC018734 179556 bp DNA linear HTG 07-JUL-2000  
DEFINITION Homo sapiens chromosome 21 clone RP11-351D2, WORKING DRAFT  
SEQUENCE, 10 unordered pieces.

AC018734

AC018734.3 GI:8569950

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo

1 (bases 1 to 179556)

Waterston,R.H.

AUTHORS The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 179556)

Waterston,R.H.

AUTHORS Direct Submission

JOURNAL Submitted (17-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

On Jun 16, 2000 this sequence version replaced gi:7023168.

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H.NH0351D02

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 170077 bases at least Q40

Consensus quality: 173263 bases at least Q30

Consensus quality: 175196 bases at least Q20

Insert size: 180000; agarose-fp

Insert size: 178656; sum-of-contigs

Quality coverage: 5.54 in Q20 bases; agarose-fp

Quality coverage: 5.64 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 10 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

-----

1 2355: contig of 2355 bp in length

2356 2455: gap of unknown length

2456 6730: contig of 4275 bp in length

6731 6830: gap of unknown length

6831 17211: contig of 10381 bp in length

17212 17311: gap of unknown length

17312 30665: contig of 13354 bp in length

30666 30765: gap of unknown length

30766 45152: contig of 14387 bp in length

45153 45252: gap of unknown length

45253 62957: contig of 17704 bp in length

62957 63056: gap of unknown length

63057 79824: contig of 16868 bp in length

79825 80024: gap of unknown length

80025 107432: contig of 27408 bp in length

107433 107532: gap of unknown length

107533 138094: contig of 30562 bp in length

138095 138194: gap of unknown length

138195 179556: contig of 41362 bp in length.

-----

Location/Qualifiers

1..179556

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/clone="RP11-351D2"

1..2355

/note="assembly\_name:Contig6"

clone\_end:SP6

## misc\_feature



```
Center project name: GUIK
Center clone name: CH230-144P21
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 199093 bases at least Q40
Consensus quality: 201840 bases at least Q30
Consensus quality: 203801 bases at least Q20
Estimated insert size: 205170; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 6 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
* 1 158307: contig of 158307 bp in length
* 158308 158407: gap of unknown length
* 158408 178882: contig of 20475 bp in length
* 178883 178982: gap of unknown length
* 178983 198357: contig of 19375 bp in length
* 198358 198457: gap of unknown length
* 198458 209332: contig of 10875 bp in length
* 209333 209432: gap of unknown length
* 209433 210436: contig of 1004 bp in length
* 210437 210536: gap of unknown length
* 210537 211897: contig of 1361 bp in length.
* Location/Qualifiers
  source
    1..211897
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-144P21"
  misc_feature
    1..1700
    /notes="wgs contig"
  gap
    158308..158407
    /estimated_length=unknown
  gap
    178883..178982
    /estimated_length=unknown
  misc_feature
    183448..184725
    /notes="wgs contig"
  gap
    198358..198457
    /estimated_length=unknown
  gap
    209333..209432
    /estimated_length=unknown
  gap
    210437..210536
    /estimated_length=unknown
ORIGIN
Query Match      87.6%; Score 18.4; DB 14; Length 211897;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTC 20
    |||||
Db 103579 TGCATGTATGGGTATCTTC 103560
    |||||
RESULT 60
AP001751
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 95/105. PRI 21-MAY-2003
DEFINITION AP001751 AL163296 BA000005
ACCESSION AP001751.1 GI:7768753
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
10830953
REFERENCE
2 (bases 1 to 340000)
AUTHORS
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuayama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
10830953
REFERENCE
2 (bases 1 to 340000)
AUTHORS
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuayama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submission
Submitted (10-Apr-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
On May 30, 2000 this sequence version replaced gi:7717412.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan.
* e-mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e-mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
160-8582, Japan,
* e-mail: nshimizu@dmb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-planck Institute for Molecular Genetics,
* Ihnestrass 73, D-14195 Berlin, Germany,
* e-mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163296: Submitted (10-Apr-2000) .
Location/Qualifiers
1..340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
1..158063
source
```

FEATURES  
source

source

source /organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP001046"  
15173. .193275  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KUD45D10"  
/clone\_lib="KU21-D Cosmid library"  
/note="Accession No. AP001047"  
192809. .209397  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="21q22.3"  
/clone="KB1216G2"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP001048"  
206177. .242765  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KUD41H11"  
/clone\_lib="KU21-D Cosmid library"  
/note="Accession No. AP001049"  
222082. .321579  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KB953G5"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP001050"  
293727. .2340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KB161B12, 3' partial"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP001051"  
1802. .1827  
/note="(TTC)A)n"  
repeat\_region /rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
complement (3874. .4044)  
/note="MERSA"  
repeat\_region /rpt\_family="DNA/MER1\_type"  
/rpt\_type=DISPERSED  
complement (4382. .4654)  
/note="L2"  
repeat\_region /rpt\_family="LINE/L2"  
/rpt\_type=DISPERSED  
4756. .4986  
/note="AluSx"  
repeat\_region /rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (5691. .6047)  
/note="MLT1A1"  
/rpt\_family="LTR/MaLR"  
/rpt\_type=DISPERSED

repeat\_region complement (11209. .11341)  
/note="MIR"  
/rpt\_family="SINE/MIR"  
/rpt\_type=DISPERSED  
repeat\_region complement (11335. .11465)  
/note="MIR"  
/rpt\_family="SINE/MIR"  
/rpt\_type=DISPERSED  
repeat\_region complement (11839. .12177)  
/note="MLT1A1"  
/rpt\_family="LTR/MaLR"  
/rpt\_type=DISPERSED  
repeat\_region complement (12280. .12538)  
/note="L2"  
/rpt\_family="LINE/L2"  
/rpt\_type=DISPERSED  
12539. .13330  
/note="(GGGGA)n"  
/rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
repeat\_region 13365. .13608  
/note="(GGGGA)n"  
/rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
repeat\_region complement (15402. .15494)  
/note="MIR"  
/rpt\_family="SINE/MIR"  
/rpt\_type=DISPERSED  
repeat\_region 15947. .16036  
/note="L2"  
/rpt\_family="LINE/L2"  
/rpt\_type=DISPERSED  
repeat\_region 16175. .16347  
/note="(TCCA)n"  
/rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
repeat\_region 16371. .16545  
/note="(TCCA)n"  
/rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
repeat\_region 16546. .16804  
/note="LIMB5"  
/rpt\_family="LINE/L1"  
/rpt\_type=DISPERSED  
repeat\_region 16805. .17026

Query Match 87.6%; Score 18.4; DB 8; Length 340000;  
Best Local Similarity 95.0%; Pred.No. 92;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATCGGTATCTTC 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 90265 TGCATGTATGAGTTATCTTC 90284

Search completed: January 27, 2006, 23:13:44  
Job time : 829.191 secs

**This Page Blank (uspto)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 981.18 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaataaagagactattcgtgcaa 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_ets.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	CS109172	Sequence
2	25	100.0	1731	CQ647508	Sequence
3	25	100.0	1734	AX608629	Sequence
4	25	100.0	21781	AE014229	Streptococcus
5	25	100.0	29072	AX602188	Sequence
6	25	100.0	167050	1 SAG766847	Streptococcus
7	25	100.0	34980	CQ655069	Sequence
8	25	100.0	34980	6 AX954529	Sequence
9	23.4	93.6	3797	1 AB027569	Streptococcus
10	20.2	80.8	2242	1 STRPHOSPHO	Streptococcus
11	20.2	80.8	2259	1 STRPEIPA	Streptococcus
12	20.2	80.8	8924	1 AE014911	Streptococcus
13	20.2	80.8	110000	1 CP000029.12	Continuation (13 o
14	20.2	80.8	153074	14 AC163395	Streptococcus
15	20.2	80.8	194538	9 AC102008	Mus muscu
16	20.2	80.8	300029	1 AE016748	Staphylococcus
17	19.8	79.2	180692	9 AC134559	Mus muscu
18	19.8	79.2	223306	9 AC125181	Mus muscu

19	19.8	79.2	233163	14	AC113904	AC113904 Rattus no
20	19.8	79.2	235700	14	AC126298	AC126298 Rattus no
21	19.8	79.2	237969	14	AC121708	AC121708 Rattus no
22	19.8	79.2	246244	14	AC074330	AC074330 Mus muscu
23	19.8	79.2	307723	14	AC094573	AC094573 Rattus no
24	19.4	77.6	63233	8	AL359504	AL359504 Human DNA
25	19.4	77.6	164485	14	AC025340	AC025340 Homo sapi
26	19.2	76.8	11962	1	AE007858	AE007858 Clontridi
27	19.2	76.8	67233	14	AC103751	AC103751 Homo sapi
28	19.2	76.8	125319	15	AC164249	AC164249 Medicago
29	19.2	76.8	138075	14	AC149834	AC149834 Zea mays
30	19.2	76.8	138943	8	AC073928	AC073928 Homo sapi
31	19.2	76.8	156618	8	AC012362	AC012362 Homo sapi
32	19.2	76.8	176602	8	AC009306	AC009306 Homo sapi
33	19.2	76.8	182536	14	AC154981	AC154981 Bos tauru
34	19.2	76.8	182547	8	AC146141	AC146141 Pan trogl
35	19.2	76.8	183491	8	CNS01DU4	Continuation (3 of
36	19.2	76.8	182026	9	AC157660	Continuation (3 of
37	19.2	76.8	194173	8	CNS01RGW	Continuation (3 of
38	19.2	76.8	259603	9	AC136716	Continuation (3 of
39	19.2	76.8	309502	14	AC152434	Continuation (3 of
40	18.8	75.2	110000	14	BX901880_2	Continuation (3 of
41	18.8	75.2	202017	5	BX323460	Continuation (3 of
42	18.8	75.2	223724	5	CR391929	Continuation (3 of
43	18.8	75.2	225849	14	AC098905	Continuation (3 of
44	18.8	75.2	228044	5	BX682552	Continuation (3 of
45	18.8	75.2	231498	14	AC120660	Continuation (3 of
46	18.6	74.4	575	10	BV515800	Continuation (3 of
47	18.6	74.4	694	10	BV552932	Continuation (3 of
48	18.6	74.4	703	10	BV529432	Continuation (3 of
49	18.6	74.4	940	10	BV521567	Continuation (3 of
50	18.6	74.4	1812	2	AY513464	Continuation (3 of
51	18.6	74.4	2461	1	AY064171	Continuation (3 of
52	18.6	74.4	3991	1	AY253327	Continuation (3 of
53	18.6	74.4	44243	8	BS000014	Continuation (3 of
54	18.6	74.4	56610	14	CR735128_4	Continuation (3 of
55	18.6	74.4	58651	14	AC017558	Continuation (3 of
56	18.6	74.4	69992	2	AC005421	Continuation (3 of
57	18.6	74.4	84055	2	AC004439	Continuation (3 of
58	18.6	74.4	91303	15	ATP2413	Continuation (3 of
59	18.6	74.4	95506	8	AC002407	Continuation (3 of
60	18.6	74.4	108958	8	AL451127	Continuation (3 of

## ALIGNMENTS

RESULT 1	CS109172	CS109172	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	CS109172	CS109172	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
DEFINITION	CS109172	CS109172	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
ACCESSION	CS109172	CS109172	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
VERSION	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
KEYWORDS	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
SOURCE	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
ORGANISM	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
REFERENCE	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
AUTHORS	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
TITLE	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
JOURNAL	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
FEATURES	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
source	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
Location/Qualifiers	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
1..25	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
/organism="synthetic construct"	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
/mol_type="unassigned DNA"	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
/db_xref="taxon:32630"	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005
/note="Oligonucleotide"	CS109172.1	GI:68148021	Sequence 3 from Patent EPI541697.	25 bp	DNA	linear	PAT 22-JUN-2005

## ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 CAAATTAAGAGACTATTCTGCAA 25
    |||||||
Db 1 CAAATTAAGAGACTATTCTGCAA 25

RESULT 2
CQ647508 1731 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 4465 from Patent WO234771.
DEFINITION CQ647508
ACCESSION CQ647508
VERSION CQ647508.1 GI:41683480
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
Nucleic acids and proteins from streptococcus groups a & b
TITLE Patent: WO 0234771-A 4465 02-MAY-2002;
JOURNAL Chiron S.p.A. (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1. .1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGCAA 25
    |||||||
Db 265 CAAATTAAGAGACTATTCTGCAA 289

RESULT 3
AX608629 1734 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 6558 from Patent WO2092818.
DEFINITION AX608629
ACCESSION AX608629
VERSION AX608629.1 GI:28404206
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
TITLE Patient: WO 02092818-A 6558 21-NOV-2002;
JOURNAL INSTITUT PASTEUR (FR) : CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
Location/Qualifiers
source
1. .1734
/organisms="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGCAA 25
    |||||||
Db 265 CAAATTAAGAGACTATTCTGCAA 289

```

RESULT 4  
AE014229  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

FEATURES  
source

gene  
CDS

gene  
CDS

gene

AE014229 21781 bp DNA linear BCT 23-SEP-2002  
Streptococcus agalactiae 2603V/R section 39 of 100 of the complete  
genome  
AE014229.1 GI:22533831

Streptococcus agalactiae 2603V/R  
Streptococcus agalactiae 2603V/R  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

1 (bases 1 to 21781)

Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,  
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,  
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,  
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,  
Madupu,K., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,  
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,  
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,  
Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,  
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.

Complete genome sequence and comparative genomic analysis of an  
emerging human pathogen, serotype V Streptococcus agalactiae  
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)  
12200547

2 (bases 1 to 21781)

Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,  
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,  
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,  
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,  
Madupu,K., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,  
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,  
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,  
Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,  
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.

Direct Submission

Submitted (18-JUL-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .21781

/organism="Streptococcus agalactiae 2603V/R"

/mol\_type="genomic DNA"

/strain="2603V/R"

/serotype="V"

/db\_xref="taxon:208435"

103. .348

/gene="SAG0813"

103. .348

/gene="SAG0813"

/notes="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAM99700.1"

/db\_xref="GI:22533832"

/translation="MNQINHCFKSIMVLGVITQMKKEFITGRNVARVMTLVCLCMV  
SNVLVLDLIDIIINEIVFCVLVASIAGLEFLCKERE"

387. .674

/gene="SAG0814"

387. .674

/gene="SAG0814"

/notes="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAM99701.1"

/db\_xref="GI:22533833"

/translation="MTYIGVPKEQKALSRGKLEKINQGGKTLMSQIMODLVGSTCM  
IYLEDGSGTHPYEILKVEDWDVQIRRNLRKSEVRIVRIHDIKELRPVKN"

719. .934

```
/gene="SAG0815"
719..934
/notes="SAG0815"
/notes="identified by match to PFAM protein family HMM
PF01381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/CI family"
/protein_id="AA099702.1"
/db_xref="GI:22533834"
/translacion="MGLIKRKLKELRARDINGNQTEAKLTGVSROTISLIERGEYTP
SVIAAKISQVFKESLENFVLSMEE"
934..1695
/gene="SAG0816"
934..1695
/gene="SAG0816"
/notes="similar to OMN1.NTL01SPL1088; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AA099703.1"
/db_xref="GI:22533835"
/translacion="MMKNFTIKQDFTFVRRIKSLIGFGFVGVGALFIIFGQDK
YLSEINIVQVFLWVSIIVITLALFSLIYLIQIQKQVFFNFVDSQSEIYRQINLR
HSQMPFVSISIVLSIVNTLFNKLAIPODSVTLPIVDLSLFLVLLGLHIYFLKVY
NIRGKINMTVAPIRLKELKNVLQDEALESNYKMGCFIDVMNLSGFIFPIYVLFPI
SFVQKVEIVITTSIHIYIILIKSLKAARHFR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/notes="similar to OMN1.NTL01SPL0446; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AA099704.1"
/db_xref="GI:22533836"
/translacion="MSKNNTTCLITETAIFAALAMALSMTPDFASWTFPSFGAIPIL
FALRRGTGYGLWGLLHLFVLSLQVFEYILAFISGLAGVFAKFRD
ALSSSTKSLALSGAILATLVRYVWHYIAGVWASYPKMGSAATLSLSVNGTA
GLTLRFVWISIIILVISYPSFPLPKK"
complement(2584..3543)
/gene="nrdf-2"
/notes="synonym: SAG0818"
complement(2584..3543)
/gene="nrdf-2"
/notes="identified by match to PFAM protein family HMM
PF00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AA099705.1"
/db_xref="GI:22533837"
/translacion="MTTYEAINNEIEDVDKSTWELKTEQFWLDRIPLSNLDLDM
RKLSAQKDLGVKVGGLTLDLTWQSETGVEATRADVTPHEEAVLNNIQPMESVHAK
SYSISFTLNTKSEIEIFWNTNNEFLQEKARIINDIYANGNALOKKVASLYLETEL
FYSGFPTPLVYLGNNKLANVAEIKLIIRDES VHGTGYGVKFGQFGLPEDEENFR
DMWYDLYOLYENKTKLYDGVGWTVEVMTFLRYNANKALMLGQDPLFPDPTAND
VNPVIMGISTGTNSNHPFPFQVGVNGYLLGSVEAMHDDDDYNGL"
complement(3746..5905)
/gene="nrdf-2"
/notes="synonym: SAG0819"
complement(3746..5905)
/gene="nrdf-2"
/notes="identified by match to PFAM protein family HMM
PF00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
/gene="SAG0820"
complement(5983..6207)
/gene="nrdf"
/notes="synonym: SAG0820"
complement(5983..6207)
/gene="nrdf"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2,
NrdfH-reodoxin"
/protein_id="AA099707.1"
/db_xref="GI:22533839"
/translacion="MAAITVFSKNNCMQCKMTKKFLDQHGADPEEINIDEKPKIEYV
KNLGSAPVIEAGNVFVSFGPSPKLKELV"
6590..6853
/gene="ptsh"
/notes="synonym: SAG0821"
6590..6853
/gene="ptsh"
/notes="identified by match to PFAM protein family HMM
PF00381"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AA099708.1"
/db_xref="GI:22533840"
/translacion="MASKDPHIVATGTHARPATLLVQTASKFASDITLDYKGKAVNL
KSLMGVMSLGVGGQADVTTISAEAGADADDAIAAETWTKEGLA"
6858..8591
/gene="ptsh"
/notes="synonym: SAG0822"
6858..8591
/gene="ptsh"
/notes="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AA099709.1"
/db_xref="GI:22533841"
/translacion="MTEMLKGIASDGVAVAKAYLLVQPLSPETVTVEDTNAEERL
DVALQSOEELSIVIREKVEAEAAVFDALHMLVSDPEMINQIKETIRAKQVNAE
TGLKEVDMFITIFEGMEDPNYQOERAADIRDVAKRVLAHLGLVKLPNPATINEESIV
IADLPSPDQAQNLKQPVKAFVNTICGRTSHSAIMARTLEIAAVLGNIDITERVQDQG
Query Match 100.0%; Score 25; DB 1; Length 21781;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CAAATTAAGAGACTATTCGTGCA 25
Db 7122 CAAATTAAGAGACTATTCGTGCA 7146
RESULT 5
AX602188
LOCUS 29072 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 117 from Patent WO02092818.
ACCESSION AX602188
```

VERSION AX602188.1 GI:28402057  
KEYWORDS Streptococcus agalactiae  
SOURCE Streptococcus agalactiae  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
REFERENCE 1  
AUTHORS Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L., Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and Kunat,F.  
TITLE Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets  
JOURNAL Patent: WO 0292818-A 117 21-NOV-2002;  
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)  
FEATURES  
source Location/Qualifiers  
1..29072  
/organism="Streptococcus agalactiae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1311"  
ORIGIN  
Query Match 100.0%; Score 25; DB 6; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAATTAAAGAGACTATTCTGTCAA 25  
|||||  
Db 12160 CAAATTAAAGAGACTATTCTGTCAA 12184  
RESULT 6  
SAG766847 167050 bp DNA linear BCT 16-APR-2005  
LOCUS Streptococcus agalactiae NEM316 complete genome, segment 5.  
DEFINITION AL766847 AL732656  
ACCESSION AL766847.1 GI:23095200  
VERSION  
KEYWORDS  
SOURCE Streptococcus agalactiae NEM316  
ORGANISM Streptococcus agalactiae NEM316  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
REFERENCE 1  
AUTHORS Glaser,P., Rusniok,C., Buchrieser,C., Chevalier,F., Frangeul,L., Msadek,T., Zouine,M., Couve,E., Lalioui,L., Poyart,C., Trieu-Cuot,P. and Kunst,P.  
TITLE Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease  
JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)  
PUBMED 12354221  
REFERENCE 2  
AUTHORS Glaser,P., Rusniok,C. and Frangeul,L.  
TITLE Direct Submission  
JOURNAL Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86  
FEATURES  
source Location/Qualifiers  
1..167050  
/organism="Streptococcus agalactiae NEM316"  
/mol\_type="genomic DNA"  
/strain="NEM316"  
/serovar="serovar III"  
/db\_xref="taxon:211110"  
join(288..293,303..1412)  
/gene="gbs0767"  
288..293  
/gene="gbs0767"  
303..1412  
/gene="gbs0767"  
/notes="Similar to riboflavin specific deaminase (diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil

reductase)"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46411.1"  
/db\_xref="GI:23095201"  
/db\_xref="GOA:Q8E660"  
/db\_xref="InterPro:IPR002125"  
/db\_xref="InterPro:IPR002734"  
/db\_xref="InterPro:IPR004794"  
/db\_xref="InterPro:IPR011549"  
/db\_xref="UniProt/TREMBL:Q8E660"  
/translation="MMKEDYMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYH  
KRGDLHAERQAIKNAEDI SGSTLYVTLEPCCHVGKOPCTEALIKSGIKVVGSL  
DPNPLVSGKIALKKECLNVEVGILREECALNERFIFHMTYKQPFYLYKAVMTLDG  
KIATKTSKWI SNESHQKURQKCSAIMVGINTVLADNFRITCRIPKGEALVRI  
VCDSQLKIPDSYLVKSARTPTWTATCSDLNAQQOITLKMGCRILIKVRPKDKLDLK  
VLMITLQEGIDSLIIEGSSLHFSALKAGIVNRLIVFIAPKIIIGLKAKTASGEGL  
DWLNQDFRVKDIELSRMDSDDVIEGKVEHYVYRNY"  
1381..1387  
/gene="gbs0768"  
1381..1387  
/gene="gbs0768"  
1393..2043  
/gene="gbs0768"  
1393..2043  
/gene="gbs0768"  
/note="Similar to ribiflavin synthase alpha chain"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46412.1"  
/db\_xref="GI:23095202"  
/db\_xref="InterPro:IPR001783"  
/db\_xref="UniProt/TREMBL:Q8E659"  
/translation="MFTGIIIEBMGQVSRIRNGIKSQLSIDAPKLVLPLRKGDSDVAVN  
GVCLTVLDKSETAFADTVMPESMMRTSLAALRLHKSVALDLALRDSRLRGHFVLGHV  
DVGKIEIKQDDIAVRFSIDAPPSIMSYIIEKGSVALDGLSTVVSFTEHSEFVSVI  
PHTMAQTNLKLVGDLNIEVDVLCKYAEKFLAPTNRNHTSSVMDMSFLSENGY"  
2046..3254  
/gene="gbs0769"  
2046..2051  
/gene="gbs0769"  
2061..3254  
/gene="gbs0769"  
/note="Similar to GTP cyclohydrazase II/  
3,4-dihydroxy-2-butanone-4-phosphate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46413.1"  
/db\_xref="GI:23095203"  
/db\_xref="GOA:Q8E658"  
/db\_xref="InterPro:IPR000422"  
/db\_xref="InterPro:IPR000926"  
/db\_xref="UniProt/TREMBL:Q8E658"  
/translation="WAFSPFKLLQDIKSGKMMVMDNRENEGDLICAAEMVTKES  
INFMAKFGKGLICLPFLSNYAEKLELAQASHNTDNHTAFTISIDHLSTSTGISABD  
RLTAKMVANDSSKAKDFRRPGHFLPLAKKEGVLARNGHTATVDLCRLAGKECGL  
CCEINAEQDSMMRKDELLAFAQHDLATATIKQLQDYRQREEGGVVREIEIQLPTQFG  
HFTAYGVSEVAVNKEHVALKGDISSEDVLCRLHSECTIGDVFHSRLCQCGQLANA  
LQOIEAERGVLVLYMRQGRGIGLNLKAYHLQEGDLTLEANLALGFEGERDYG  
GHLLD"  
3255..3739  
/gene="gbs0770"  
3255..3260  
/gene="gbs0770"  
3269..3739  
/gene="gbs0770"  
/note="Similar to riboflavin synthase complex beta chain"  
/codon\_start=1  
/transl\_table=11

```
/product="Unknown"
/protein_id="CAD46414.1"
/db_xref="GI:23095204"
/db_xref="GOA:Q8E657"
/db_xref="InterPro:IPR002180"
/db_xref="UniProt/Swiss-Prot:Q8E657"
/translation="MTIIEQLVANEMKIGIVVSRFENLITSKLKSGAVDGLLRHGVG
EBDIDIVVPGAFIPIYMAKWMALYKDYDAIICLGVIKGTSDHYDVYVCNEVTKGIGH
LNSQSDIPHIFGVLTIDNIEQALIERAGTKAGNGYDCAISAIEWNLDKLRER"
complement(3814..5315)
/gene="lysS"
complement(3814..5304)
/gene="lysS"
/codon_start=1
/transl_table=11
/product="lysyl-tRNA synthetase"
/protein_id="CAD46415.1"
/db_xref="GI:23095205"
/db_xref="GOA:Q8E656"
/db_xref="InterPro:IPR002313"
/db_xref="InterPro:IPR004364"
/db_xref="InterPro:IPR004365"
/db_xref="InterPro:IPR006195"
/db_xref="InterPro:IPR008994"
/db_xref="UniProt/Swiss-Prot:Q8E656"
/translation="MSNOHIELENDQIVRRKMAALTEQIDPFGKRFERTATSGQL
NEKYADKSEDLHDIEETATIAGRLMKRGKGVGAHIOREGOIQIYVRKDSVGEF
NYEIFKADGLDFVEGQVMRTDMGELSIKATHITLSKALAPLEKFGHGLTDITI
YRKHLDLISNRDSFDRFVTRKIISEIRFMSNGFLEVETPLVNEAGSARFPI
THNQAOIDMWLAIATLHLKRLIVGMERVYIGIRIFRNEGDAHNPFTSIEAYQ
AVADYQIDMLTGTIOHVTYKVGDPINYOQTEIKINEPKRHHMVDVAKIEIGID
FWKEMTELAQAQKKNVPLEKHFTTVGHIINARFEFVEDTLIOPTVFGHPVVS
PLAKNDTPRFTDRLELFIPTYKEYANATELNDPIDQLSRBEAQAASEKLGDEATG
VDYDYVALEIYMPPTGGIGIDRLCMLLDTITIRDVLLFPFMK"
complement(5310..5315)
/gene="lysS"
5467..6411
/gene="gbs0772"
5467..5472
/gene="gbs0772"
5479..6381
/gene="gbs0772"
/notes="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46416.1"
/db_xref="GI:23095206"
/db_xref="GOA:Q8E655"
/db_xref="InterPro:IPR005833"
/db_xref="InterPro:IPR005834"
/db_xref="InterPro:IPR006439"
/db_xref="UniProt/TrEMBL:Q8E655"
/translations="MITSIVFDVDDTIYDQAPYIAVEKCFDFDMSAINQAIYRFR
HYSIDGPRVWAGETWTFYFRFWRKETLLEFGVEIDEATGIYFOBIYSHLENIITM
LDERWTLFLSKNVPWGIITNGPTEHQLKKVKKGLGLYDVPKRVISQATGFOKPK
EKGIIFNLAAQSPDNFSTLLYVGDSYDNDINGAFNGGHHSHWFNHRGRSLKPKIKVPY
DVAIDNPEQLFGAVKVLFDLPDNKRFIDINDKSNPVMELGNLGLMMAERLLESNMS
VDKVIILLRLTAQEKVLRMKYAR"
6390..6411
/gene="gbs0772"
complement(6417..7069)
/gene="gbs0773"
complement(6417..7058)
/gene="gbs0773"
/notes="Similar to putative phosphoglycerate mutase"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46417.1"
/db_xref="GI:23095207"
/db_xref="GOA:Q8E654"
/db_xref="InterPro:IPR001345"
```

```
/db_xref="UniProt/TrEMBL:Q8E654"
/translation="MKFYLVRHGKTQMNLEGRFOGANGDSPLEBAIEBELEBLGQYLS
SIHFDAVYSSDLGRADTVNINDANSCPKHEIHYTPQLREWALGTLEGCKIATMOAIY
PRMTAFYQNPLQPKHDMFGAESLYQTHRVESFLRSLASKNYDKVLIYVGHGANLTAS
IRSLGQYGSLSHYKDKLDNASLTIIETHDFKDPNCLTWNDKSYLRQEVKMTM"
complement(7064..7069)
/gene="gbs0773"
complement(7080..7553)
/gene="gbs0774"
complement(7080..7553)
/gene="gbs0774"
/notes="Similar to transcriptional regulator"
/codon_start=1
/transl_table=11
Query Match 100.0%; Score 25; DB 1; Length 167050;
Best Local Similarity 100.0%; Pred. No. 2.1; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25
Db 79926 CAAATTAAAGAGACTATTCGTGCAA 79950

RESULT 7
CO655069 349980 bp DNA linear PAT 02-FEB-2004
Sequence 12026 from Patent WO0234771.
DEFINITION CO655069
ACCESSION CO655069
VERSION CO655069.1 GI:41687946
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12026 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1..349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/notes="seq 10967 to big: splitted in
seq 10.967, from 0.000.001 to 0.349.980
seq 12.025, from 0.300.001 to 0.649.980
seq 12.026, from 0.600.001 to 0.949.980
seq 12.027, from 0.900.001 to 1.249.980
seq 12.028, from 1.200.001 to 1.549.980
seq 12.029, from 1.500.001 to 1.849.980
seq 12.030, from 1.800.001 to 2.149.980
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25
Db 191703 CAAATTAAAGAGACTATTCGTGCAA 191727

RESULT 8
AX954529 349980 bp DNA linear PAT 08-JAN-2004
LOCUS AX954529
DEFINITION Sequence 1375 from Patent WO03093306.
ACCESSION AX954529
VERSION AX954529.1 GI:40783902
KEYWORDS
```

```

SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
AUTHORS Telford,J., Masignani,V., margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a b
JOURNAL Patent: WO 03093306-A 1375 13-NOV-2003;
Chiron SRL (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source Location/Qualifiers
1. .349980
/organism="Streptococcus pyogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1314"
/no/seqs="seq 1373 to long, original length: 2.160.266
replaced by
seq 1373: from 0.000.001 to 0.349.980
seq 1374: from 0.300.001 to 0.649.980
seq 1375: from 0.600.001 to 0.949.980
seq 1376: from 0.900.001 to 1.249.980
seq 1377: from 1.200.001 to 1.549.980
seq 1378: from 1.500.001 to 1.849.980
seq 1379: from 1.800.001 to 2.149.980
seq 1380: from 2.100.001 to 2.160.266"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGCGAA 25
|||||
Db 220490 CAAATTAAAGAGACTATTCTGCGAA 220514

RESULT 9
AB027569 3797 bp DNA linear BCT 23-JUL-2003
LOCUS Streptococcus bovis ptsH, ptsI genes for histidine containing
DEFINITION protein, phosphoenolpyruvate-protein phosphotransferase, complete
cds.
ACCESSION AB027569
VERSION AB027569.3 GI:6723275
KEYWORDS phosphoenolpyruvate-protein phosphotransferase; histidine
containing protein.
SOURCE Streptococcus bovis
ORGANISM Streptococcus bovis
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
AUTHORS Aasanuma,N. and Hino,T.
TITLE Molecular characterization of Hpr and related enzymes, and
regulation of Hpr phosphorylation in the ruminal bacterium
Streptococcus bovis
JOURNAL Arch. Microbiol. 179 (3), 205-213 (2003)
PUBMED 12610726
AUTHORS Aasanuma,N. and Hino,T.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-1999) Narito Aasanuma, Meiji University,
Department of Agriculture; Higashimita, Tama-ku, Kawasaki, Kanagawa
214-8571, Japan (E-mail: aasanuma@isc.meiji.ac.jp).
COMMENT Tel: +81-44-934-7825 (ex.7825), Fax: +81-44-934-7825)
On Jan 20, 2000 this sequence version replaced gi:5706360.
FEATURES
source Location/Qualifiers
1. .3797
/organism="Streptococcus bovis"
/mol_type="genomic DNA"
/strain="JB1"
/db_xref="taxon:1315"
1779. .2042
/gene="ptsH"

gene

```



JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14434-14439 (2002)
PUBMED	12397186
REFERENCE	2 (bases 1 to 8924)
AUTHORS	Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J., Carson,M.B., Prineaux,C., Tian,R., Kenton,S., Jia,H., Lin,S., Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and Ferretti,J.J.
TITLE	Direct Submission
JOURNAL	Submitted (09-JUL-2002) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd., Oklahoma City, OK 73104, USA
FEATURES	Location/Qualifiers
source	1..8924
	/organism="Streptococcus mutans UA159"
	/mol_type="genomic DNA"
	/strain="UA159"
gene	/db_xref="taxon:210007"
	365..1162
CDS	365..1162
	/locus_tag="SMU.673"
	365..1162
	/locus_tag="SMU.673"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAN58407.1"
	/db_xref="GI:24377052"
	/translation="MWISINLVFLFFYISFIGNLWETVYCSLKDKQFYVRFGLAGPY CPVGFATVTLFKPOENLPILFFGLIIATVFYMAACFLKVFHKLWDYSQE SNGIRAPRISLFWGIVLFKFIQPSVNMQLLAKNWNWALGIALFMFADFMTV SGLAQFORAALSLEKMSAEGDKIOAYLREIWNLEEDTDFYLSIRLFHLQIEKMR RXGIQPFPHQRMRLNYHKLKADPFAVAAVHKLASVKNKDKKR"
gene	1328..1591
	/gene="ptsH"
	1328..1591
CDS	1328..1591
	/locus_tag="SMU.674"
	1328..1591
	/gene="ptsH"
	/locus_tag="SMU.674"
	/note="Best Blastp Hit: sp P45596 PTHP_STRMU PHOSPHOCARRIER PROTEIN HPR (HISTIDINE-CONTAINING PROTEIN)
	>gi 482380 pir  A44562 phosphotransferase system phosphohistidine-containing protein - Streptococcus mutans
	>gi 310627 gb AA91092.1 (L15191)
	phosphoenolpyruvate:sugar phosphotransferase system Hpr [Streptococcus mutans]"
	/codon_start=1
	/transl_table=11
	/product="phosphoenolpyruvate:sugar phosphotransferase system Hpr"
	/protein_id="AAN58408.1"
	/db_xref="GI:24377053"
gene	/translation="WASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGAVNL KSIIMGVMSIGVGQADVTITABGADADDARAINETMTKEGLA"
	1596..3329
CDS	1596..3329
	/locus_tag="SMU.675"
	1596..3329
	/locus_tag="SMU.675"
	/note="Best Blastp Hit: sp P45595 PTL_STRMU PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I)
	>gi 310628 gb AA91093.1 (L15191)
	phosphoenolpyruvate:sugar phosphotransferase system enzyme I [Streptococcus mutans]"
	/codon_start=1
	/transl_table=11
	/product="phosphoenolpyruvate:sugar phosphotransferase system enzyme I, pts system EI component"
	/protein_id="AAN58409.1"
	/db_xref="GI:24377054"
	/translation="WTEMLUKGIAASDVAVAKAYLLVQPDLTFTETVSVTDTQAEARL DAALKEASQNELSIRQKAVDTLGEAAAVFDALHVLADPEMIGIKETIRTKYNAE SALKVEQNFVLTFENQMDPNYQEAARDIRDAKRVLAHLGLVLPNPAITSEESIV IADHLTPSDAOLDANYKAVFTNIGGRTSHSAIWAARTLEIAAVLGTNDITSRVKNGD IYAVNGITQGVINPTEDQIAEFKAAGETYAKQKAEWALLKDAETVTADGKHFEAAN

IGTKDVEGVNNGAEAVGLYRTEFLYMDSDPFTEDQEYAEYKAVLEGNGKPYVVR TMDIGGDKELPYFDLPKEMNDFLGRALRISISSETGNOMPTQLRALRASVHGQLRI NFMVALLNEFRKAKGILEERKANLKAEGVAVSDDIQVGINIEIPAAWLADQFAKEV DFFSINGNDLLQYTMADRMNEQYSYLYQPNPSILRLVDHVHVAHAHEGKAGNCGE MAGDOTAVPLLVGLGDFEFSMSATSVLRTSLMKKLDPTDKQELAQRALTSCATMEEV LELEKEVIDFD"	3559..4986	gene
/gene="gapN"		
/locus_tag="SMU.676"		
3559..4986		CDS
/gene="gapN"		
/locus_tag="SMU.676"		
/EC_number="1.2.1.9"		
/note="Best Blastp Hit: sp Q59931 GAPN_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (NADP+)) (TRIOSEPHOSPHATE DEHYDROGENASE)		
>gi 2117515 pir  A57151 glyceraldehyde-3-phosphate dehydrogenase (NADP+) (EC 1.2.1.9) - Streptococcus mutans		
>gi 5822059 pdb 1EUH A Chain A, Apo Form Of A NADp Dependent Aldehyde dehydrogenase From Streptococcus Mutans		
>gi 5822060 pdb 1EUH B Chain B, Apo Form Of A NADp Dependent Aldehyde dehydrogenase From Streptococcus Mutans		
>gi 5822061 pdb 1EUH C Chain C, Apo Form Of A NADp Dependent Aldehyde dehydrogenase From Streptococcus Mutans		
>gi 5822062 pdb 1EUH D Chain D, Apo Form Of A NADp Dependent Aldehyde dehydrogenase From Streptococcus Mutans		
>gi 5822470 pdb 2EUH A Chain A, Holo Form Of A NADp Dependent Aldehyde dehydrogenase Complex With NADp+		
>gi 5822471 pdb 2EUH B Chain B, Holo Form Of A NADp Dependent Aldehyde dehydrogenase Complex With NADp+		
>gi 5822472 pdb 2EUH C Ch"		
/codon_start=1		
/transl_table=11		
/product="NADP-dependent glyceraldehyde-3-phosphate dehydrogenase"		
/protein_id="AAN58410.1"		
/db_xref="GI:24377055"		
/translation="WTKQKNYNGEWKLSNEIKIYPASGAELGSPVPMSTERVDY VYAKKAQAPARSLYSIERAAYLHKVADILMRDKKIGAVLSKEVAKGYSVSEVV RTAETINYAABEGLMEGEVLEGGSFEAASKKTAVVRPEVPLVLAISPFPYNLA QKAPALIAGNVIAFPKPTQGSISGLLLAEFAEAGLPAGVFNITIRGSGSIGYIV EHQAVFINFTSGTIGICRGMAGMRPIMLELGGKSAIVLEADADLTAKNIITAGA FGYSQRCATVKRVLMVESVADELVEKREKVALTICNPEDDADITPLDTPKSADYV EALINDKNGAALATEIRKGNLICPILFKVYTTDMRLANEEPPGVLPIRVTSVE BAISINSEYGLQASIFTNDFPRAFGIAEQLEVGTVHNNKTQKGTDNFPFLGAKKS GAGIQGVKYSIEAMTTVKSVVFDIK"		
5657..6049		gene
/locus_tag="SMU.677"		
5657..6049		CDS
/locus_tag="SMU.677"		
/note="Best Blastp Hit: pir  B69970 transcription regulator MerR family homolog yraB - Bacillus subtilis		
>gi 2108269 emb CAA63468.1 (X92868) mercuric resistance operon regulatory protein [Bacillus subtilis]		
>gi 2635146 emb CAB14642.1 (Z99117) similar to transcriptional regulator (MerR family) [Bacillus subtilis]"		
/codon_start=1		
/transl_table=11		
/product="putative transcriptional regulator (MerR family)"		
/protein_id="AAN58411.1"		
/db_xref="GI:24377056"		
/translation="MNKKVSEITDISADTIYRIGLLPRITRNSGVDRFTFERI EILFVRCFRKAGMSVEALIBYISLIEEGCTERERLLLTQERDMDDRIYELNQAR BELNAYKIENYENIIQKHQKLFTESEEN"		
6049..6891		gene
/locus_tag="SMU.678"		
6049..6891		CDS
/locus_tag="SMU.678"		
/note="Best Blastp Hit: pir  A72308 oxidoreductase,		



aldo/keto reductase family - Thermotoga maritima (strain MSB8) >gi|4981536|gb|AAD36074.1|AE001762.1 (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]  
/codon\_start=1  
/transl\_table=11  
/product="putative oxidoreductase, aldo/keto reductase family"  
/protein\_id="AAN58412.1"  
/db\_xref="GI:24377057"  
/translation="METVVLNSGVKMPILGFGVQPDLEECERVVTEAIKVGYYLLD TAQVYGNESAVGNKSLGAREEFITTKWISNAGYEKAKASDESRLKQTDYVD LVLIHOPGDYGTGSALEYKSKAIGVSNFYFDRLVLDLASFNEIVPMVNQVET HYVNOIEAKKYSSEYGVOLSGPFAEKGNNFTNPTLTATGEKYGKSAAGVGLRYL IORGVVVIPKTVHQRWQNFDFVGFNLSDDDMADILKLDQGESLFINHYDPTVKFL ANK"

gene 6901. .7767  
/locus\_tag="SMU.679"  
CDS 6901. .7767  
/locus\_tag="SMU.679"  
/notes="Best Blastp Hit: pir||A72308 oxidoreductase, aldo/keto reductase family - Thermotoga maritima (strain MSB8) >gi|4981536|gb|AAD36074.1|AE001762.1 (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]"  
/codon\_start=1  
/transl\_table=11  
/product="putative oxidoreductase, aldo/keto reductase family"  
/protein\_id="AAN58413.1"  
/db\_xref="GI:24377058"  
/translation="MEYITLTGVMKPMGFGVQVTPDPEICKESVLNAIKTGYRLFD TAAYGVNEKALGEAHEAIEQGLVTRNELFITSKLVQDMYNEAAKKGIERSLEKGL DYLDLVLHQAQDRYFGAALAEAYQAQKRAIGVSNFYFNILTNFCTSVISVPAVN QIELHPFFQEEALETFARYYGVVPEAPLGGGVDAFNNAILQVESAHQKTVQGM LKWNQGVVVVPIPKSTHKRIENFAINDFSLTDDMTKIASLDMGYGDTRTKHFDPN FVRSVLNVKIID"

Query Match 80.8%; Score 20.2; DB 1; Length 8924;  
Best Local Similarity 88.0%; Pred. No. 3.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGTCGCAA 25  
Db 1860 CAGATTAAGAGACAATTGTCACAA 1884

RESULT 13  
CP000029.12/c  
WPCOMENT  
Sequence split into 27 fragments LOCUS CP000029 Accession CP000029  
Fragment Name Begin End  
CP000029\_00 1 110000  
CP000029\_01 100001 210000  
CP000029\_02 200001 310000  
CP000029\_03 300001 410000  
CP000029\_04 400001 510000  
CP000029\_05 500001 610000  
CP000029\_06 600001 710000  
CP000029\_07 700001 810000  
CP000029\_08 800001 910000  
CP000029\_09 900001 1010000  
CP000029\_10 1000001 1100000  
CP000029\_11 1100001 1210000  
CP000029\_12 1200001 1310000  
CP000029\_13 1300001 1410000  
CP000029\_14 1400001 1510000  
CP000029\_15 1500001 1610000  
CP000029\_16 1600001 1710000  
CP000029\_17 1700001 1810000  
CP000029\_18 1800001 1910000  
CP000029\_19 1900001 2010000  
CP000029\_20 2000001 2110000

CP000029\_21 2100001 2210000  
CP000029\_22 2200001 2310000  
CP000029\_23 2300001 2410000  
CP000029\_24 2400001 2510000  
CP000029\_25 2500001 2610000  
CP000029\_26 2600001 2616530  
Continuation (13 of 27) of CP000029 from base 1200001 (CP000029 Staphylococcus epidermi.

Query Match 80.8%; Score 20.2; DB 1; Length 110000;  
Best Local Similarity 88.0%; Pred. No. 2.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGTCGCAA 25  
Db 79426 CAAATTAAGAGAAAATTGTCGCAA 79402

RESULT 14  
AC163395/c  
LOCUS  
DEFINITION  
AC163395  
AC163395.5 GI:70721483  
HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 153074)  
Birren,B., Nuebaum,C. and Lander,E.  
Mus musculus chromosome 7, clone RP24-251J14  
Unpublished  
2 (bases 1 to 153074)

Birren,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (10-JUN-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 153074)  
Birren,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,

O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**TITLE**  
**JOURNAL**  
 Direct Submission  
 Submitted (12-JUL-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

**COMMENT**  
 On Jul 12, 2005 this sequence version replaced gi:68268145.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information

Center project name: L33069

Center clone name: 251\_J\_14

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 79735: contig of 79735 bp in length  
 \* 79736 79835: gap of unknown length  
 \* 79836 107825: contig of 27990 bp in length  
 \* 107826 107925: gap of unknown length  
 \* 107926 119432: contig of 11507 bp in length  
 \* 119433 119532: gap of unknown length  
 \* 119533 153074: contig of 33542 bp in length.

#### FEATURES

source

```
1..153074
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosomes="7"
    /map="7"
    /clone="RP24-251J14"
    /clone_lib="RP24-24 Male Mouse BAC"
79736..79835
    /estimated_length=unknown
107826..107925
    /estimated_length=unknown
119433..119532
    /estimated_length=unknown
```

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 153074;  
 Best Local Similarity 88.0%; Pred. No. 2.1e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CAATTAAGAGACTATTCGTGCAA 25  
 |||||  
 DB 23616 CAATAAAGAGAAATATCGTCAA 23592  
 |||||  
 RESULT 15  
 AC102008/c  
 LOCUS  
 DEFINITION  
 AC102008  
 AC102008.9 GI:45825240  
 HTG  
 Mus musculus (house mouse)

#### ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

**REFERENCE**  
 1 (bases 1 to 194538)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 7, clone RP24-502L8

**TITLE**  
**JOURNAL**

Unpublished

2 (bases 1 to 194538)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**Direct Submission**

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 194538)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hagos, B., Hagopian, D., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**Direct Submission**

Submitted (13-FEB-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 194538)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hagos, B., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,



JOURNAL	Submitted (05-NOV-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China
FEATURES	Location/Qualifiers
source	1..300029
	/organism="Staphylococcus epidermidis ATCC 12228"
	/mol_type="genomic DNA"
	/strain="ATCC 12228"
	/db_xref="ATCC:12228"
gene	/db_xref="taxon:176280"
CDS	complement(137..1123)
	/locus_tag="SE1168"
	complement(137..1123)
	/locus_tag="SE1168"
	/codon_start=1
	/transl_table=11
	/product="thioredoxin reductase-like protein"
	/protein_id="AAO04767.1"
	/db_xref="GI:27315632"
	/translation="MQTIESIIIGGPGCLSAAIEQKKGIETLVIEKGNVVESTYNY
	PTHQTFPSSDKLSGDIPFVEDSKPRNQALVYREVVKHQNLNHPFEVLTVKK
	INNKFAITTTKGVIECKYLTVAHYGGQHTLEAGGELPQVHYFKGAHPFNQNVV
	IIGKNSAVDALEKAGANVTLYRGEQYPAIKPWLFPNFSVLNHEKITWFEFA
	TWKTIDHSVTYKQGLIEIDNDYFAMIGYHPDYDFLKTIGIDHTNEXGTAPVYN
	RETFTNVENCYAGVIAAGNDANTIFIENGKYHGGVITQSILTKKQPLET"
	complement(1256..2674)
gene	/locus_tag="SE1169"
CDS	complement(1256..2674)
	/locus_tag="SE1169"
	/codon_start=1
	/transl_table=11
	/product="elastin binding protein"
	/protein_id="AAO04768.1"
	/db_xref="GI:27315633"
	/translation="MIEIIFKMEGMIVSNNNFKDFEKNRQISINPDEQTELKEDDKT
	NENKEADQSNLSNNOQFPFPAQRKRRETATNQSKQDDKHQKNSDAKTTEG
	SLDDRYDAQLQQHDKSQONKTEKQSDNEMKGDAAI VNGTSESPHKSKSTON
	RPGRQAQQKRSESTQSKPSTNKKKAATGAGIAGAGVAGAAETSKRHNNKKDQD
	SKHSNDEKSVKNDQDKQSKKGAAGVAGAAAGVAGAAHNNQKHNNEKNS
	NONQNYNDQEGKKGKGMKILLPIAAILLIGAIAIFGGWALNNHNSKSDNKKIAN
	QSKDQSKDQSDNKKDKSDNKKDKSDNADDDSDNSSNPATSTNNNDYV
	ANNNGNYTNQONNANQSNQATQGOQSHTVYQENLYRIAIFYYGEGTQANVVK
	IKRANGLSSNNIHNQGLVLPQ"
gene	complement(2795..4174)
CDS	/locus_tag="SE1170"
	complement(2795..4174)
	/locus_tag="SE1170"
	/codon_start=1
	/transl_table=11
	/product="probable ATP-dependent DNA helicase RecQ"
	/protein_id="AAO04769.1"
	/db_xref="GI:27315634"
	/translation="MLKNLKEWFGDFEKGQEBIIDSVLHGRHTLGIILPTGSKSL
	CYQLPTYILEKPTLII SPLISLMDQVMQTKLGETHVACVHSGMDVEKRDNIKRLS
	KSFRTYLSPEYILQPHNFKYIAHINFGILVLEAHCLEMGYDFRPHYALVGKIIHFH
	NSATVALATAPSOLETDLSHLSIKLNVOQSNRRENI SIQHYNFNDDEGKIEWLL
	PFLNSGPTIIVYSKKRCLELAQMIYDSGYLTGICYHGDLSYQERTVQQQFPNDIP
	VIVATSAFGKKDIDRTVYIHFHLSSSPSNYLQIQRAGRDGKQSAISLFPQDPSF
	ILETLFTDITDITDITPFTGPHLPDEKELISTLNTQFTFSQLKDI FHOSYQRRL
	GYMRWMTNLDQCRKYLLEFPGEHPQSPKQCCQDSRLPIQILNRKKVKRKLSPN
	EKLYNLFET"
gene	complement(4161..5123)
CDS	/locus_tag="SE1171"
	complement(4161..5123)
	/locus_tag="SE1171"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAO04770.1"
	/db_xref="GI:27315635"
	/translation="MYNILSYAMSHAFNYKTNKSIYNILTGKSKHQTFDASSQQLLS
	LYLHLPNKKYSTFEQFLOKDDFKKSIQVKIHPQVYDSLTQTFSCIQLLIHTLSHTR
	KESNTFIPIVQNTYIQORVQLYHOVIESNOVSNITDEIYLLFENLNKKNHTFLHY
	LOGYESMYTRQI SLIEGI PQSELPEREMNELIDILNQLKDS TKYPILSQAILSLPL
	LNTYLYGQKLKSLNLKEIAQLONKVLNTIEDHILEMYIKGYLIDYTLFLTKKIDILE
	FINYYQKHGRGERLKFYKEHFDWTYFQIKLVIVGIERGDLIAER"
gene	5232..5480
CDS	/locus_tag="SE1172"
	5232..5480
	/locus_tag="SE1172"
	/codon_start=1
	/transl_table=11
	/product="ferredoxin"
	/protein_id="AAO04771.1"
	/db_xref="GI:27315636"
	/translation="MAKYTIIVDMDTCIACGACGAAPDIYDDDEGIAYVILDDNQGT
	AEVPELYEDMEDALEGCTDSIKIIEDEPFGDALKFE"
	complement(5572..5676)
gene	/locus_tag="SE1173"
CDS	complement(5572..5676)
	/locus_tag="SE1173"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAO04772.1"
	/db_xref="GI:27315637"
	/translation="MSSLVPNKQPMNHKQIFWGSIHILIADYFIVYG"
	complement(5885..6424)
gene	/locus_tag="SE1174"
CDS	complement(5885..6424)
	/locus_tag="SE1174"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAO04773.1"
	/db_xref="GI:27315638"
	/translation="WQONKRLITITMLSAVAVLTFTIKFPLFPPIPYLTLDPSDVPTL
	LATLFLSPIAGIIVALIKNLINLFLNIGDPGVPANFLAGVSLSSYYVYRKRKNR
	SLIYGLITGTIVMTIVLSILNYFVLLPLYGMIFNLGDLVNNVIVISGVIPFNLRIG
	IIIGIIFVLLFRRLRHIIK"
	complement(7016..8782)
gene	/locus_tag="SE1175"
CDS	complement(7016..8782)
	/locus_tag="SE1175"
	/codon_start=1
	/transl_table=11
	/product="staphylococcal respiratory response protein
	SrrB"
	/protein_id="AAO04774.1"
	/db_xref="GI:27315639"
	/translation="MNRLSNVIKMLWTIILIVATVLVLLSVSLITFIQYFTQETEN
	AIKEDAKRISSLVKSHNKTIAIQNSQTLIDGPGGLIIMKMSKDPDSSYKDTKKQ
	MFNEIKKSTKFKYKEGETQNTITIKNGNSQSYLLGLYPMKQAQGAQSHYSGVFI
	YKDLKSTDTNNAITIIITAIITIASTIAPFLSNRI TKPLRQLTKQAQKVSSEG
	YSQISTVATKDEIGDLSAFNNMVEIOEHKAISSKNIRDTLLNSMVGVLGINNO
	RIEILSNKADDIRHDDPFKSEIEQIEQIEATFESQQNEYLELEINTRYVYFISYID
	RIQNGRSGIYVMVIRDMTNEHNLQMKDFIANVSHELURTPISLQGYTESLVDGIVT
	EPDEIRLSLAIVLDESKRLNRLNVLNARMDAEGLSVEKELQPIQHLDRKMSKYR
	MQSELGTMTFDSNDEQLWNVDMDQVLTNLIDNATRYTQAGDSIKISIDSDS
	FNILITDTGTGIAPEHLKQVDFYKVDAAARKKGQCTGLGLFCIKMIIEBGGRID
	VESBLGKGTSTFIILRPKSKQIS"
	complement(8766..9491)
gene	/locus_tag="SE1176"
CDS	complement(8766..9491)
	/locus_tag="SE1176"
	/codon_start=1
	/transl_table=11
	/product="staphylococcal respiratory response protein
	SrrA"
	/protein_id="AAO04775.1"
	/db_xref="GI:27315640"
	/translation="WTNEILIVDDDEDIRRLRLKLYLRESFEIHEARDKEAYELALE
	NNYACILDLMLPEWDGIEVASKLRHEKDTPIIMLTAKGESTNRVEGESGADDYIVK
	PFSPREVVLRVALLRRQTQANTQSEPHARDIIEFSLHVIDNDARHVLADDOQVNL
	PKFELVLYLAKTRKPRFDRQLKVBWYHBYFGDLRTVDTHTVKKLRKLNKVSDDAA
	QMIIQTVMGVGYKFEVKSDEPTK"
	complement(9625..10362)

```
CDS      /locus tag="SE1177"
          complement(9625..10362)
          /locus_tag="SE1177"
          /codon_start=1
          /transl_table=11
          /product="ribosomal large subunit pseudouridine synthase
          B"
          /protein_id="AA004776.1"
          /db_xref="GI:27315641"
          /translation="MSKEQERLQKRIANSGYTSRRKAENVLITEGRVKNQGVTVTELGS
          KWPMSDLVVDGKIIEQDKIYILFYKPTQVITSVSDRGRVVDYFDLLESRIYPV
          GRLDYTSGLLTNDGEFTNLTPHYHIKKKYAKGLYMLRBEIKQLEKGBLEL
          GYTPAQKIKQDKENITLVEITISEGRNQVRMFYFGHQVKNLTRIQFGPIDL
          KGLNAGEGRVLPHEVKILRLHLAENGK"
          complement(10355..10897)
          /locus_tag="SE1178"
          /locus_tag="SE1178"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA004777.1"
          /db_xref="GI:27315642"
          /translation="MDNIAILEALLYTSGDEGLEQKQIIDIILNINQLELDLVSKYHS
          HGLTIORYGSTVLTTKETSTVIEQLVKEKSKMLSQAMETLSIIAYNQPLTRGDI
          EMTRGINSQAVKTLIARGLVEAKVDHRSRSHLITTDLFNVFGIENLDALPTTEED
          EAEMDEFFSNLVNQKESNE"
          complement(10890..11690)
          /locus_tag="SE1179"
          /locus_tag="SE1179"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA004778.1"
          /db_xref="GI:27315643"
          /translation="MTSCYNNFVKYALTMQFDIEVDMYEVKLDNFNGPLDILLHLIQ
          KYEIDYIDPMKALTSQYQYVHAMQLEINVAEYLVAESLLMIKSKLLLPOTSIE
          EDIEDREDRLVCRLEIYQYKEYTEILNTMKERDLYTKHPTDLTHLETRESWDPN
          QTIDITELIVAYQQRVNEVLPKSVIEIKETFTIQQAQVTERLKQHESFNFFSL
          FTFPEVQVVTTHFLAILLEMSKSGIIVNIKQTKQFDIDDIIRGVNYSIG"

Query Match      80.8%; Score 20.2; DB 1; Length 300029;
Best Local Similarity 88.0%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CAAATTAAGAGACATTCGTGCAA 25
        |||||
        179897 CAAATTAAGAGAAATTCGTGCAA 179873

RESULT 17
AC134559/c
LOCUS      AC134559.5 180692 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-332C4 from chromosome 8, complete
sequence.
ACCESSION AC134559
VERSION   AC134559.5 GI:31746699
KEYWORDS  HTG.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
           1 (bases 1 to 180692)
REFERENCE 1 The sequence of Mus musculus BAC clone RP24-332C4
AUTHORS   Isak.A.
TITLE     Unpublished (2001)
JOURNAL   Unpublished (2001)
REFERENCE 2 (bases 1 to 180692)
AUTHORS   Wilson,R.
TITLE     Sequencing of Mus musculus
JOURNAL   Unpublished (2001)
REFERENCE 3 (bases 1 to 180692)
```

```
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park
        Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 180692)
AUTHORS  Wilson,R.K.
TITLE    Direct Submission
JOURNAL  Submitted (31-MAY-2003) Genome Sequencing Center, 4444 Forest Park
        Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 180692)
AUTHORS  Wilson,R.K.
TITLE    Direct Submission
JOURNAL  Submitted (15-JUN-2003) Genome Sequencing Center, 4444 Forest Park
        Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 180692)
AUTHORS  Wilson,R.
TITLE    Direct Submission
JOURNAL  Submitted (27-NOV-2003) Department of Genetics, Washington
        University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
        On Jun 15, 2003 this sequence version replaced gi:31249897.
COMMENT  ----- Genome Center
        Center: Washington University Genome Sequencing Center
        Center code: WUGSC
        Web site: http://genome.wustl.edu
        Contact: submissions@watson.wustl.edu
        ----- Summary Statistics
        -----
        Center project name: M_BB0332C04
        -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compression and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:  
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (http://www.chori.org) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone. This clone is overlapped by AC125181.

FEATURES  
source  
1..180692  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="8"  
/map="8"  
/clone="RP24-332C4"  
/clone\_lib="RPCI-24"  
3651..3810  
/rpt\_family="B4"  
8421..8518  
/rpt\_family="Alu"  
9813..9926  
/rpt\_family="B4"  
12341..12502  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region

repeat\_region /rpt\_family="B2"  
15504. .15570  
repeat\_region /rpt\_family="ERV1"  
16825. .17053  
repeat\_region /rpt\_family="B2"  
17236. .17289  
repeat\_region /rpt\_family="L1"  
17354. .17757  
repeat\_region /rpt\_family="MaLR"  
18008. .18168  
repeat\_region /rpt\_family="B4"  
19179. .19336  
repeat\_region /rpt\_family="B4"  
22885. .23056  
repeat\_region /rpt\_family="B4"  
23404. .23749  
repeat\_region /rpt\_family="L1"  
23811. .23901  
repeat\_region /rpt\_family="MIR"  
23929. .24118  
repeat\_region /rpt\_family="L1"  
24400. .24592  
repeat\_region /rpt\_family="B4"  
25087. .25192  
repeat\_region /rpt\_family="Alu"  
25088. .25233  
repeat\_region /rpt\_family="B4"  
28351. .28546  
repeat\_region /rpt\_family="ERVK"  
30305. .31233  
repeat\_region /rpt\_family="L1"  
31379. .31463  
repeat\_region /rpt\_family="MIR"  
32867. .32941  
repeat\_region /rpt\_family="L1"  
33285. .33333  
repeat\_region /rpt\_family="Alu"  
33334. .33429  
repeat\_region /rpt\_family="L1"  
33480. .33671  
repeat\_region /rpt\_family="B2"  
33707. .33888  
repeat\_region /rpt\_family="MER1\_type"  
33878. .34150  
repeat\_region /rpt\_family="MaLR"  
34657. .34794  
repeat\_region /rpt\_family="Alu"  
34663. .34826  
repeat\_region /rpt\_family="B4"  
35656. .35799  
repeat\_region /rpt\_family="Alu"  
36988. .37362  
repeat\_region /rpt\_family="MaLR"  
37616. .37803  
repeat\_region /rpt\_family="B2"  
37970. .38064  
repeat\_region /rpt\_family="Alu"  
38235. .38596  
repeat\_region /rpt\_family="L1"  
38865. .38987  
repeat\_region /rpt\_family="Alu"  
41406. .41521  
repeat\_region /rpt\_family="B4"  
42792. .42940  
repeat\_region /rpt\_family="B4"  
43582. .44132  
repeat\_region /rpt\_family="L1"  
44171. .45313  
repeat\_region /rpt\_family="L1"  
45384. .45868  
repeat\_region /rpt\_family="L1"  
47036. .47108  
repeat\_region /rpt\_family="ID"

repeat\_region 52478. .52602  
/rpt\_family="L1"  
repeat\_region 52616. .52810  
/rpt\_family="B2"  
repeat\_region 52898. .53722  
/rpt\_family="L1"  
repeat\_region 55180. .55382  
/rpt\_family="B2"  
repeat\_region 55342. .55497  
/rpt\_family="B4"  
repeat\_region 56728. .56848  
/rpt\_family="L1"  
repeat\_region 56849. .57044  
/rpt\_family="B2"  
repeat\_region 57045. .57126  
/rpt\_family="L1"  
repeat\_region 57325. .57510  
/rpt\_family="B2"  
repeat\_region 57815. .57929  
/rpt\_family="Alu"  
repeat\_region 57993. .58231  
/rpt\_family="L1"  
repeat\_region 58315. .58522  
/rpt\_family="B2"  
repeat\_region 59688. .59871  
/rpt\_family="B2"  
repeat\_region 60098. .60283  
/rpt\_family="ERVK"  
repeat\_region 60321. .60395  
/rpt\_family="ID"  
repeat\_region 60409. .60535  
/rpt\_family="B4"  
repeat\_region 61091. .61416  
/rpt\_family="MaLR"

Query Match 79.2%; Score 19.8; DB 9; Length 180692;  
Best Local Similarity 91.3%; Pred. No. 3e+02; Mismatches 0; Gaps 0;  
Matches 21; Conservative 0; Indels 0;

QY 2 AAATTAAGAGACTATTCTGTGCA 24  
|||  
Db 125231 AATTAAAGATACTATTCTGTGCA 125209  
|||  
|||

RESULT 18  
AC125181/c 223306 bp DNA linear ROD 11-NOV-2003  
LOCUS Mus musculus BAC clone RP23-324L6 from chromosome 8, complete  
DEFINITION sequence.  
ACCESSION AC125181  
VERSION AC125181.4 GI:28927902  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 223306)  
AUTHORS Swearingen-Shahid,S., Kozlowicz,A., Doebber,A., Shahid,S.,  
Haglund,K. and Haakenson,W.  
The sequence of Mus musculus BAC clone RP23-324L6  
Unpublished (2001)  
REFERENCE 2 (bases 1 to 223306)  
AUTHORS Wilson,R.  
Sequencing of Mus musculus  
Unpublished (2001)  
REFERENCE 3 (bases 1 to 223306)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
Direct Submission  
TITLE Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
JOURNAL Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 223306)  
AUTHORS McPherson,J.D. and Waterston,R.H.



```

repeat_region      /rpt_family="MER1_type"
58372..58931
/rpt_family="ERVK"
repeat_region      58932..59080
/rpt_family="Alu"
59081..59388
/rpt_family="ERVK"
repeat_region      59595..59923
/rpt_family="L1"
60655..61177
/rpt_family="L1"
repeat_region      61370..61844
/rpt_family="L1"
repeat_region      61845..61991
/rpt_family="Alu"
repeat_region      61992..62394
/rpt_family="L1"
repeat_region      62504..62551
/rpt_family="ERV1"
repeat_region      63152..69230
/rpt_family="ID"
repeat_region      70669..70846
/rpt_family="RMR15"
repeat_region      73521..73615
/rpt_family="Alu"
repeat_region      73651..73849
/rpt_family="B2"

```

```

Query Match      79.2%; Score 19.8; DB 9; Length 223306;
Best Local Similarity 91.3%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 AAATTAAGAGACTATTCGTGCA 24
|| ||||| ||||| ||||| |||||
Db 2721 AATTTAAGATACACTATTCGTGCA 2699

```

```

RESULT 19
AC113904
LOCUS      Rattus norvegicus clone CH230-318P21, *** SEQUENCING IN PROGRESS
DEFINITION *** 2 unrounded clones.
AC113904
VERSION    AC113904.5 GI:25072640
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
            1 (bases 1 to 233163)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,P., Haves,A., Henderson,N., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

```

```

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensunewa,L., Louised,H., Lozad,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murph,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nair,L.,
Nwakoelameh,O., Okwuonu,G., Olarnpunaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ben,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,I., Rojas,A., Rose,M., Rose,R., Ruiz,S.O.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 233163)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233163)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23194968.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNAV
Center clone name: CH230-318P21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 209814 bases at least Q40
Consensus quality: 211069 bases at least Q30
Consensus quality: 211999 bases at least Q20
Estimated insert size: 220261; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

```



```

* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 223189: contig of 223189 bp in length
* 223190 223289: gap of unknown length
* 223290 233163: contig of 9874 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..233163
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10116"
                /clone="CH230-318P21"
            1..2826
                /note="wgs_end_extension"
                /clone_end="T7"
            7180..8127
                /note="clone_boundary"
                /clone_end="T7"
                site:
            end_sequence=BZ141951"
            223190..223289
                /estimated_length=unknown
ORIGIN
Query Match      79.2%; Score 19.8; DB 14; Length 233163;
Best Local Similarity 91.3%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATTAAGACGACTATTCTGTCGCA 25
    |||||
Db 13272 ATTAAGACGACTAATTGTGCA 13294
    |||||

RESULT 20
AC126298/c
LOCUS
DEFINITION
AC126298 Rattus norvegicus clone CH230-31B17, *** SEQUENCING IN PROGRESS
***, 9 unordered pieces.
AC126298 3 GI:23196133
AC126298.3 HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 235700)
Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cres,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebraegorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

```

```

Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorusshewa,L., Louleeged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norrie,S.,
Nwaokemelemb,O., Okwuonu,G., Olarnpunaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shan,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,
Sheed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 235700)
Worley,K.C.
Direct Submission
Submitted (05-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 235700)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21702798.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYMB
Center Clone name: CH230-31B17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 161627 bases at least Q40
Consensus quality: 168027 bases at least Q30
Consensus quality: 172221 bases at least Q20
Estimated insert size: 192265; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 10630: contig of 10630 bp in length  
\* 10631 10730: gap of unknown length  
\* 10731 22411: contig of 21381 bp in length  
\* 22411 22412: gap of unknown length  
\* 22412 22527: contig of 1064 bp in length  
\* 22527 22537: gap of unknown length  
\* 22537 22638: contig of 1009 bp in length  
\* 22638 22648: gap of unknown length  
\* 22648 22808: contig of 1599 bp in length  
\* 22808 22818: gap of unknown length  
\* 22818 22928: contig of 1045 bp in length  
\* 22928 22938: gap of unknown length  
\* 22938 23073: contig of 1395 bp in length  
\* 23073 23083: gap of unknown length  
\* 23083 23136: contig of 2313 bp in length  
\* 23136 23236: gap of unknown length  
\* 23236 23570: contig of 2464 bp in length.  
\* 23570 23571: Location/Qualifiers

## FEATURES

source  
1. 235700  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-31B17"  
1. 1234  
/notes="wgs contig"  
10631..10730  
/estimated\_length=unknown  
30002..31205  
/notes="wgs contig"  
42447..43559  
/notes="wgs contig"  
56213..57746  
/notes="wgs contig"  
103185..104507  
/notes="wgs contig"  
121152..121288  
/notes="clone boundary"  
clone end:T7  
site:ECORI  
end sequence:BH319296"  
220901..222807  
/notes="wgs\_end\_extension"  
clone end:T7"  
224112..224211  
/estimated\_length=unknown  
225276..225375  
/estimated\_length=unknown  
226385..226484  
/estimated\_length=unknown  
228084..228183  
/estimated\_length=unknown  
229229..229328  
/estimated\_length=unknown  
230724..230823  
/estimated\_length=unknown  
23137..23236  
/estimated\_length=unknown

## ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 235700;  
Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AATTAAAGAGACTATTCGTCAA 25  
|||||  
Db 139577 AATTAAAGAGACTAATTGTGCAA 139555

RESULT 21

AC121708/c

. LOCUS

237969 bp DNA linear

HTG 21-SEP-2002

## DEFINITION

AC121708  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM

## ORGANISM

REFERENCE  
AUTHORS

Rattus norvegicus clone CH230-190A9, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 3 unordered pieces.

AC121708

GI:23265128

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 237969)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, P., Garcia, A., Garner, T., Garza, M.,  
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerrero, I., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, H., L., Loule, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwackeleme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D.,  
Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

## TITLE

Unpublished  
JOURNAL  
REFERENCE  
2 (bases 1 to 237969)  
Worley, K. C.

## AUTHORS

Direct Submission  
TITLE  
JOURNAL  
Submitted (21-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 237969)

## REFERENCE

Rat Genome Sequencing Consortium.

## AUTHORS

Direct Submission  
TITLE  
JOURNAL  
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
 On Sep 21, 2002 this sequence version replaced gi:21909182.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GYGS  
 Center clone name: CH230-190A9  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 169159 bases at least Q40  
 Consensus quality: 174233 bases at least Q30  
 Consensus quality: 177883 bases at least Q20  
 Estimated insert size: 191644; sum-of-contigs estimation  
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 233860: contig of 233860 bp in length  
 \* 233861 233960: gap of unknown length  
 \* 233961 235605: contig of 1645 bp in length  
 \* 235606 235705: gap of unknown length  
 \* 235706 237969: contig of 2264 bp in length.

## FEATURES

source	1. .237969 Location/Qualifiers /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-190A9"
misc_feature	1. .1164 /note="wgs end_extension clone_end:T7"
misc_feature	1498. .3143 /note="wgs_end_extension clone_end:T7"
misc_feature	complement(4467. .5177) /note="clone_boundary clone_end:T7 site:EcoRI end sequence:BH362335"
misc_feature	68167. .70215 /note="wgs contig"
misc_feature	104654. .105710 /note="wgs contig"
misc_feature	108284. .108038 /note="wgs contig"
misc_feature	200400. .201597 /note="wgs contig"
misc_feature	complement(233121. .233860) /note="clone_boundary clone_end:Sp6 site:EcoRI end sequence:BH362336"
gap	233861. .233960 /estimated length=unknown
gap	235606. .235705

/estimated\_length=unknown

## ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 237969;  
 Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 AATTAAAGAGACTATTCTGTGCAA 25  
 |||||  
 Db 46221 AATTAAAGAGACTAATTGTGCAA 46199  
 |||||  
 RESULT 22  
 AC074330/c  
 LOCUS  
 DEFINITION Mus musculus strain C57BL/6J chromosome 12 clone RP23-439L22,  
 WORKING DRAFT SEQUENCE, 35 unordered pieces.  
 AC074330 246244 bp DNA linear HTG 27-JUN-2002  
 AC074330.2 GI:11225644  
 HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_CANCELLED.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 Smith,D.R.  
 1 (bases 1 to 246244)  
 Genome Therapeutics Corporation Sequencing Center: Mouse Genome  
 Sequence Data  
 Unpublished  
 2 (bases 1 to 246244)  
 Smith,D.R.  
 Direct Submission  
 Submitted (23-AUG-2000) Genome Therapeutics Corporation, 100 Beaver  
 Street, Waltham, MA 02453, USA  
 On Nov 19, 2000 this sequence version replaced gi:9885992.  
 ----- Genome Center  
 Center: Genome Therapeutics Corporation  
 Center code: GTC  
 Web site: http://www.genomecorp.com/  
 Contact: gtc-seqcenter@genomecorp.com  
 ----- Project Information  
 Center project name: mg038  
 ----- Summary Statistics  
 Sequencing vector: N/A  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 990315  
 Consensus quality: 223772 bases at least Q40  
 Consensus quality: 228225 bases at least Q30  
 Consensus quality: 230837 bases at least Q20  
 Insert size: 242943; sum-of-contigs  
 Quality coverage: 5.0x in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1003: contig of 1003 bp in length  
 \* 1004 1103: gap of unknown length  
 \* 1104 2262: contig of 1159 bp in length  
 \* 2263 2362: gap of unknown length  
 \* 2363 3423: contig of 1061 bp in length  
 \* 3424 3523: gap of unknown length  
 \* 3524 4714: contig of 1191 bp in length  
 \* 4715 4814: gap of unknown length  
 \* 4815 5945: contig of 1131 bp in length  
 \* 5946 6046: gap of unknown length  
 \* 6046 7182: contig of 1137 bp in length  
 \* 7183 7282: gap of unknown length

*	7283	8431:	contig	of 1149	bp in length
*	8432	8531:	gap	of unknown	length
*	8532	9679:	contig	of 1148	bp in length
*	9680	9779:	gap	of unknown	length
*	9780	10902:	contig	of 1123	bp in length
*	10903	11002:	gap	of unknown	length
*	11003	12371:	contig	of 1269	bp in length
*	12272	12371:	gap	of unknown	length
*	12372	13579:	contig	of 1208	bp in length
*	13580	13679:	gap	of unknown	length
*	13680	14974:	contig	of 1295	bp in length
*	14975	15074:	gap	of unknown	length
*	15075	16120:	contig	of 1046	bp in length
*	16121	16220:	gap	of unknown	length
*	16221	17554:	contig	of 1334	bp in length
*	17555	17654:	gap	of unknown	length
*	17655	18714:	contig	of 1060	bp in length
*	18715	18814:	gap	of unknown	length
*	18815	20213:	contig	of 1399	bp in length
*	20214	20313:	gap	of unknown	length
*	20314	21457:	contig	of 1144	bp in length
*	21458	21557:	gap	of unknown	length
*	21558	22702:	contig	of 1145	bp in length
*	22703	22802:	gap	of unknown	length
*	22803	24679:	contig	of 1877	bp in length
*	24680	24779:	gap	of unknown	length
*	24780	26082:	contig	of 1303	bp in length
*	26083	26182:	gap	of unknown	length
*	26183	27553:	contig	of 1371	bp in length
*	27554	27653:	gap	of unknown	length
*	27654	28960:	contig	of 1307	bp in length
*	28961	29060:	gap	of unknown	length
*	29061	30604:	contig	of 1544	bp in length
*	30605	30704:	gap	of unknown	length
*	30705	32438:	contig	of 1734	bp in length
*	32439	32538:	gap	of unknown	length
*	32539	34239:	contig	of 1701	bp in length
*	34240	34339:	gap	of unknown	length
*	34340	35795:	contig	of 1456	bp in length
*	35796	35895:	gap	of unknown	length
*	35896	37708:	contig	of 1813	bp in length
*	37709	37808:	gap	of unknown	length
*	37809	38864:	contig	of 1056	bp in length
*	38865	38964:	gap	of unknown	length
*	38965	41365:	contig	of 2401	bp in length
*	41366	41465:	gap	of unknown	length
*	41466	43774:	contig	of 2309	bp in length
*	43775	43874:	gap	of unknown	length
*	43875	46518:	contig	of 2644	bp in length
*	46519	46618:	gap	of unknown	length
*	46619	49941:	contig	of 3323	bp in length
*	49942	50041:	gap	of unknown	length
*	50042	101223:	contig	of 51082	bp in length
*	101124	101223:	gap	of unknown	length
*	101224	162347:	contig	of 61124	bp in length
*	162348	162447:	gap	of unknown	length
*	162448	246244:	contig	of 83797	bp in length

misc_feature	/estimated_length=unknown 2363. .3423 /note="assembly_name:Contig94"
gap	3424. .3523 /estimated_length=unknown
misc_feature	3524. .4714 /note="assembly_name:Contig97"
gap	4715. .4814 /estimated_length=unknown
misc_feature	4815. .5945 /note="assembly_name:Contig942"
gap	5946. .6045 /estimated_length=unknown
misc_feature	6046. .7182 /note="assembly_name:Contig943"
gap	7183. .7282 /estimated_length=unknown
misc_feature	7283. .8431 /note="assembly_name:Contig946"
gap	8432. .8531 /estimated_length=unknown
misc_feature	8532. .9679 /note="assembly_name:Contig947"
gap	9680. .9779 /estimated_length=unknown
misc_feature	9780. .10902 /note="assembly_name:Contig948"
gap	10903. .11002 /estimated_length=unknown
misc_feature	11003. .12771 /note="assembly_name:Contig952"
gap	12772. .12371 /estimated_length=unknown
misc_feature	12372. .13579 /note="assembly_name:Contig953"
gap	13580. .13679 /estimated_length=unknown
misc_feature	13680. .14974 /note="assembly_name:Contig955"
gap	14975. .15074 /estimated_length=unknown
misc_feature	15075. .16120 /note="assembly_name:Contig961"
gap	16121. .16220 /estimated_length=unknown
misc_feature	16221. .17554 /note="assembly_name:Contig965"
gap	17555. .17654 /estimated_length=unknown
misc_feature	17655. .18714 /note="assembly_name:Contig975"
gap	18715. .18814 /estimated_length=unknown
misc_feature	18815. .20213 /note="assembly_name:Contig979"
gap	20214. .20313 /estimated_length=unknown
misc_feature	20314. .2157 /note="assembly_name:Contig981"
gap	21458. .21557 /estimated_length=unknown
misc_feature	21558. .22702 /note="assembly_name:Contig987"
gap	22703. .22802 /estimated_length=unknown
misc_feature	22803. .24679 /note="assembly_name:Contig988"
gap	24680. .24779

Query Match	79.2%	Score 19.8;	DB 14;	Length 246244;
Best Local Similarity	91.3%	Pred. No. 2.8e+02;		
Matches 21; Conservative		0; Mismatches 2;	Indels 0;	



```

end_sequence: BH309389"
275881. .277825
/note="wgs_end_extension
clone end: 77"
278039. .281071
/note="wgs_end_extension
clone end: 77"
281072. .281171
/estimated_length=unknown
282967. .283066
/estimated_length=unknown

ORIGIN
Query Match 79.2%; Score 19.8; DB 14; Length 307723;
Best Local Similarity 91.3%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATTAAGACAGACTATTCGTGCA 25
|||||
Db 151393 AATTAAGACAGACTAATTCGTGCA 151415
|||||

RESULT 24
AL359504
LOCUS
DEFINITION
Human DNA sequence from clone RP11-486G15 on chromosome 1 Contains
the 5' end of the PRKACB gene for cAMP-dependent catalytic protein
kinase beta, a thioredoxin 2 (TXN2) pseudogene and a CpG island,
complete sequence.
AL359504
AL359504
VERSION
AL359504.29 GI:20068411
KEYWORDS
HTG; CpG island; kinase; PRKACB; thioredoxin; TXN2.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 63233)
Van Hellmond, Z.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:18873412.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP11-486G15 is from the library RPI1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
source
1. .63233
/organism="Homo sapiens"

```

---

```

/mol_type="genomic DNA"
/db_xref="taxon.9606"
/chromosome="1"
/clone="RP11-486G15"
/clone_lib="RPCI-11.2"
2000
/note="Clone right end: RP4-552O12"
join(10256..10565,AL450063.14:50139..50200,
AL450063.14:53162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:85115..85279,AL450063.14:106142..109460)
/genes="PRKACB"
/locus_tag="RP11-82H13.1-001"
join(10256..10565,AL450063.14:50139..50200,
AL450063.14:53162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:85115..85279,AL450063.14:106142..109460)
/genes="PRKACB"
/locus_tag="RP11-82H13.1-001"
/product="protein kinase, cAMP-dependent, catalytic, beta"
/note="match: ESTs: Em:AA610029.1 Em:AA815281.1
Em:AA833749.1 Em:AL028253.1 Em:AI076037.1 Em:AI357420.1
Em:AI589447.1 Em:AL596920.1 Em:AUI29973.1 Em:BS622570.1
Em:BQ100900.1 Em:BQ101211.1 Em:BQ230651.1 Em:BQ962828.1
Em:BU194335.1 Em:BUS21071.1 Em:CA866319.1 Em:CD657149.1
Em:CF136689.1
match: CDNAs: Em:AL050208.1 Em:BC035058.1 Em:J02647.1
Em:M34181.1"
join(10338..10565,AL450063.14:50139..50200,
AL450063.14:53162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:75444..76401)
/genes="PRKACB"
/locus_tag="RP11-82H13.1-002"
join(10338..10565,AL450063.14:50139..50200,
AL450063.14:53162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:75444..76401)
/genes="PRKACB"
/locus_tag="RP11-82H13.1-002"
/product="protein kinase, cAMP-dependent, catalytic, beta"
/note="match: ESTs: Em:BF695747.1 Em:BX335182.1
Em:BX369882.1
match: CDNAs: Em:BC016285.1"
join(10520..10565,AL450063.14:50139..50200,
AL450063.14:53162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:75444..75452)
/genes="PRKACB"
/locus_tag="RP11-82H13.1-002"
/standard_name="OTTHUMP0000011664"
/note="match: proteins: Tr:Q96B09"
/codon_start=1
/product="protein kinase, cAMP-dependent, catalytic, beta"
/protein_id="CAI14540.1"
/db_xref="GI:55959225"
/db_xref="GOA:Q96B09"
/db_xref="InterPro:IPR000719"
/db_xref="InterPro:IPR001245"
/db_xref="InterPro:IPR002290"
/db_xref="InterPro:IPR008271"
/db_xref="UniProt/TrEMBL:Q96B09"
/translation="MGNAATAKGVSEVESFEFLAKAKEDFLKKWENTONNAGLEDFF
ERKTLTGSGRGNLVXGKATEQYANKILDKQVKLQKIEHTLNKRILQAVNFP
FLVRLEYAFKNSNLNLYMWQYVPGGEMFSLRRIGRFSPPHARFTAAQILVLFELHS
LDLIYRDLKPNLLIDHQYIQVTFDFAKRVKGRVTLTCLGTFPEYLAPEILUSKGYNK
AVDMWALGLIYEMAAGPPFPADQPIQIYEKIVSGKNF"
join(10520..10565,AL450063.14:50139..50200,

```

```

AL450063.14:531162..53290,AL450063.14:54999..55097,
AL450063.14:56062..56144,AL450063.14:67578..67704,
AL450063.14:68691..68786,AL450063.14:73645..73767,
AL450063.14:85115..85279,AL450063.14:106142..106267)
/gene="PRKACB"
/locus_tag="RP11-82H13.1-002"
/standard_names="OTHUMP0000011663"
/note="match: proteinins: Sw:P05131 Sw:P05383 Tr:P22694"
/codon_start=1
/product="protein kinase, cAMP-dependent, catalytic, beta"
/protein_id="Ca14541.1"
/db_xref="GI:55959226"
/translation="MGNATARKSEVESVKEFLAKAKEDFLKKWENTQNNAGLEDF
ERKKTGLTGFGRVNLKHYKATEQYYAMKILDKQVVKLQKIEHTLNKRILQAVNFP
FLVRLEYAFKNSLYMYEYVPGGEMFSLRRIGRFSEPHARFYAAOVLTFEYLHS
LDLIYRDLKPNLLIDHOGYIQVTFPGAKRVKGTWLTCTPEYLAPELLISKGYNK
AVDWALGVLIYEMAGGTPFFPADQPIQYIKVSGKVRFPFHSFSDLKDLRLNLQV
DLTRFGNLKGVSDIKTHKWFATTDWIAIYQKVEAPFIPKFRSGDTSNPFDDYEEER
DIRVITBKCAKEGFEF"
17935..18427
/locus_tag="RP11-486G15.1-001"
/pseudo
17935..18427
/locus_tag="RP11-486G15.1-001"
/note="match: proteins: Sw:P97493 Sw:P97615 Sw:Q95108
Sw:Q99757 Tr:AA05576 Tr:BAB22037 Tr:BAB27267"
/pseudo
/codon_start=1
/product="thioredoxin 2 (TXN2) pseudogene"
61234
/note="Clone_left_end: RP11-82H13"

gene
CDS
misc_feature
ORIGIN

Query Match 77.6%; Score 19.4; DB 8; Length 63233;
Best Local Similarity 95.2%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTG 22
|||||
Db 20566 AAATTAAGAGACTATTCGTG 20586

RESULT 25
AC025340/C
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-742B6 map 1, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
AC025340
ACCESSION AC025340.2 GI:7321608
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 164485)
REFERENCE
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguski, L., Boukhalil, B., Brown, A., Burkett, G.,
Campopiano, A., Castelle, A., Choe, Y., Colangelo, M., Collins, S.,
Collumore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, G., Locke, K., Liu, G., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

```

```

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 24, 2000 this sequence version replaced gi:7210048.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6992
Center Clone name: 742_B_6
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153825 bases at least Q40
Consensus quality: 159684 bases at least Q30
Consensus quality: 161668 bases at least Q20
Insert size: 152000; agarose-Ep
Insert size: 162785; sum-of-contents
Quality coverage: 4.6 in Q20 bases; agarose-ep
Quality coverage: 4.3 in Q20 bases; sum-of-contents
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2576: contig of 2576 bp in length
* 2577 2676: gap of 100 bp
* 2677 5906: contig of 3230 bp in length
* 5907 6006: gap of 100 bp
* 6007 9486: contig of 3480 bp in length
* 9487 9586: gap of 100 bp
* 9587 12060: contig of 2474 bp in length
* 12061 12160: gap of 100 bp
* 12161 18546: contig of 6386 bp in length
* 18547 18647: gap of 100 bp
* 18647 23010: contig of 4363 bp in length
* 23010 29919: gap of 100 bp
* 29920 30019: contig of 6810 bp in length
* 30020 38748: contig of 8729 bp in length
* 38749 38849: gap of 100 bp
* 38849 46149: contig of 7301 bp in length
* 46150 53642: gap of 100 bp
* 53643 53742: contig of 7393 bp in length
* 53743 63925: contig of 10183 bp in length
* 63925 64025: gap of 100 bp
* 64026 73519: contig of 9494 bp in length
* 73519 73619: gap of 100 bp
* 73620 87188: contig of 13569 bp in length
* 87189 87288: gap of 100 bp
* 87289 100090: contig of 12802 bp in length
* 100091 100190: gap of 100 bp
* 100191 111124: contig of 10934 bp in length
* 111124 111224: gap of 100 bp

```

TITLE  
JOURNAL

## COMMENT

```

* 111225 123063: contig of 11839 bp in length
* 123064 123163: gap of 100 bp
* 123164 141217: contig of 18054 bp in length
* 141218 141317: gap of 100 bp
* 141318 164485: contig of 23168 bp in length.
FEATURES
    source
        1..164485
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="1"
            /clone="RP11-742B6"
            /clone_lib="RPC1-11 Human Male BAC"
            1..2576
                /note="assembly_fragment"
                2577..2676
                    /estimated_length=100
                2677..5906
                    /note="assembly_fragment"
                5907..6006
                    /estimated_length=100
                6007..9485
                    /note="assembly_fragment"
                9487..9586
                    /estimated_length=100
                9587..12060
                    /note="assembly_fragment"
                12061..12160
                    /estimated_length=100
                12161..18546
                    /note="assembly_fragment"
                18547..18646
                    /estimated_length=100
                18647..23009
                    /note="assembly_fragment"
                    vector_side:left
                23010..23109
                    /estimated_length=100
                23110..29919
                    /note="assembly_fragment"
                29920..30019
                    /estimated_length=100
                30020..38748
                    /note="assembly_fragment"
                38749..38848
                    /estimated_length=100
                38849..46149
                    /note="assembly_fragment"
                46150..46249
                    /estimated_length=100
                46250..53642
                    /note="assembly_fragment"
                53643..53742
                    /estimated_length=100
                53743..63925
                    /note="assembly_fragment"
                63926..64025
                    /estimated_length=100
                64026..73519
                    /note="assembly_fragment"
                73520..73619
                    /estimated_length=100
                73620..87188
                    /note="assembly_fragment"
                87189..87288
                    /estimated_length=100
                87289..100090
                    /note="assembly_fragment"
                100091..100190
                    /estimated_length=100
                100191..111124
                    /note="assembly_fragment"
                    111125..111224
                        /estimated_length=100
                    111225..123063
                        /note="assembly_fragment"
                123064..123163
                    /estimated_length=100
                123164..141217
                    /note="assembly_fragment"
                141218..141317
                    /estimated_length=100
                141318..164485
                    /note="assembly_fragment"
                    clone_end:SP6
                    vector_side:right
            ORIGIN
                Query Match      77.6%; Score 19.4; DB 14; Length 164485;
                Best Local Similarity 95.2%; Pred. No. 4.5e+02;
                Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
                QY      2 AAATTAAGAGACTATTCTG 22
                        |||||
                Db      26102 AAATTAAGAGACTATTCTG 26082
                        |||||
            RESULT 26
            AE007858      11962 bp      DNA      linear      BCT 01-JUN-2004
            LOCUS      Clostridium acetobutylicum ATCC 824 section 346 of 356 of the
            DEFINITION      complete genome
            ACCESSION      AE007858 AE001437
            VERSION      AE007858.1 GI:15026716
            KEYWORDS
            SOURCE      Clostridium acetobutylicum ATCC 824
            ORGANISM      Clostridium acetobutylicum ATCC 824
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
            1 (bases 1 to 11962)
            Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
            Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
            J. Bacteriol. 183 (16), 4823-4838 (2001)
            11466286
            2 (bases 1 to 11962)
            Childress,D., Zeng,Q. and Smith,D.R.
            Direct Submission
            Submitted (24-JUL-2001) GRC Sequencing Center Production,
            Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
            Location/Qualifiers
            1..11962
            /organism="Clostridium acetobutylicum ATCC 824"
            /mol_type="genomic DNA"
            /strain="ATCC 824"
            /db_xref="ATCC:824"
            /db_xref="taxon:272562"
            268..783
            /gene="CAC3617"
            268..783
            /gene="CAC3617"
            /codon_start=1
            /transl_table=11
            /product="Uncharacterized membrane protein, YHAG
            B. subtilis homolog"
            /protein_id="AAK81540.1"
            /db_xref="GI:15026717"
            /translation="MRMNLKKLIINSLFLAVGVNLNQITPPILFGMKPFDLSAMLFII
            ILLNDYKTCISTGVAGLAAVTFPGQLPNIIDRIVTTSLVFIALRPFKDKIND
            KIHMTITTVGTIISGSVFLGSAIIIVGLPASFKALFITVVLPAIIAIVGTIIFVA

```





```

/gene="CAC3628"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATPase component"
/protein_id="AAK81551.1"
/db_xref="GI:15026728"
/translation="MAEENKSDVLEVKLVNKKYFNVGHNAITLKAVDDVSFTIKKGETL
GLVGSCKGKTCGRVIGLQVQATEGEVNVYGVNHSURGAKKFKSKAQMIFQDPY
ACLNPRMTVGDIIAGDIDVHGLYKGEETKKIYEMLDHVLNHRHSAKFPHEFGGQR

```

```

Query Match          76.8%; Score 19.2; DB 1; Length 11962;
Best Local Similarity 87.5%; Pred. No. 9 2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2  AAATTAAGAGACTATTCGTGCAA 25
        |||||  |||  |||||  |||||
Db      1652 AAATTAAGATAATTCCTGCAA 1675

```

```

RESULT 27
AC103751/c
LOCUS   Homo sapiens chromosome 15 clone RP11-536E9 map 15, LOW-PASS
DEFINITION
SEQUENCE SMPILING.
AC103751
VERSION AC103751.2 GI:22004476
KEYWORDS HTG; HTGS PHASE0.
SOURCE   Homo sapiens

```

```

ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominoidea; Homo.
REFERENCE 1 (bases 1 to 67233)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 15, clone RP11-536E9
JOURNAL   Unpublished

```

```

REFERENCE 2 (bases 1 to 67233)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
           Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

REFERENCE 3 (bases 1 to 67233)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
           Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
           Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
           Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
           Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
           Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
           McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,

```

```

Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

```

```

Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 30, 2002 this sequence version replaced gi:17149636.
All repeats were identified using RepeatMasker:

```

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21702
Center clone name: 536_E_9

```

```

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

* 1
* 647 746: contig of 646 bp in length
* 747 1446: gap of 100 bp
* 1447 1446: contig of 700 bp in length
* 1547 2252: contig of 706 bp in length
* 2253 2352: gap of 100 bp
* 2353 3042: contig of 690 bp in length
* 3043 3142: gap of 100 bp
* 3143 3839: contig of 697 bp in length
* 3840 3939: gap of 100 bp
* 3940 4616: contig of 677 bp in length
* 4617 4716: gap of 100 bp
* 4717 5422: contig of 706 bp in length
* 5423 5522: gap of 100 bp
* 5523 6232: contig of 710 bp in length
* 6233 6332: gap of 100 bp
* 6333 7043: contig of 711 bp in length
* 7044 7143: gap of 100 bp
* 7144 7855: contig of 712 bp in length
* 7856 7955: gap of 100 bp
* 7956 8652: contig of 697 bp in length
* 8653 8752: gap of 100 bp
* 8753 9462: contig of 710 bp in length
* 9463 9562: gap of 100 bp
* 9563 10260: contig of 698 bp in length
* 10261 10360: gap of 100 bp
* 10361 10441: contig of 681 bp in length
* 11042 11141: gap of 100 bp
* 11142 11848: contig of 707 bp in length
* 11849 11948: gap of 100 bp
* 11949 12653: contig of 705 bp in length
* 12654 12753: gap of 100 bp
* 12754 13439: contig of 686 bp in length
* 13440 13539: gap of 100 bp
* 13540 14259: contig of 720 bp in length
* 14260 14359: gap of 100 bp
* 14360 15077: contig of 718 bp in length
* 15078 15177: gap of 100 bp

```

\* 15178 15883: contig of 706 bp in length  
 \* 15884 15983: gap of 100 bp  
 \* 15984 16690: contig of 707 bp in length  
 \* 16690 16790: gap of 100 bp  
 \* 16790 17500: contig of 710 bp in length  
 \* 17500 17600: gap of 100 bp  
 \* 17600 18297: contig of 697 bp in length  
 \* 18297 18397: gap of 100 bp  
 \* 18397 19091: contig of 694 bp in length  
 \* 19091 19191: gap of 100 bp  
 \* 19191 19883: contig of 692 bp in length  
 \* 19883 19984: gap of 100 bp  
 \* 19984 20670: contig of 687 bp in length  
 \* 20670 20770: gap of 100 bp  
 \* 20770 21476: contig of 705 bp in length  
 \* 21476 21575: gap of 100 bp  
 \* 21575 22289: contig of 714 bp in length  
 \* 22289 22390: gap of 100 bp  
 \* 22390 23088: contig of 699 bp in length  
 \* 23088 23189: gap of 100 bp  
 \* 23189 23895: contig of 707 bp in length  
 \* 23895 23956: gap of 100 bp  
 \* 23956 24722: contig of 727 bp in length  
 \* 24722 24823: gap of 100 bp  
 \* 24823 25529: contig of 707 bp in length  
 \* 25529 25629: gap of 100 bp  
 \* 25629 26338: contig of 709 bp in length  
 \* 26338 26438: gap of 100 bp  
 \* 26438 27151: contig of 713 bp in length  
 \* 27151 27251: gap of 100 bp  
 \* 27251 27956: contig of 705 bp in length  
 \* 27956 28056: gap of 100 bp  
 \* 28056 28757: contig of 699 bp in length  
 \* 28757 28853: gap of 100 bp  
 \* 28853 29554: contig of 699 bp in length  
 \* 29554 30366: contig of 712 bp in length  
 \* 30366 30466: gap of 100 bp  
 \* 30466 31168: contig of 702 bp in length  
 \* 31168 31268: gap of 100 bp  
 \* 31268 31985: contig of 717 bp in length  
 \* 31985 32084: contig of 719 bp in length  
 \* 32084 32904: gap of 100 bp  
 \* 32904 33606: contig of 702 bp in length  
 \* 33606 33707: gap of 100 bp  
 \* 33707 34393: contig of 687 bp in length  
 \* 34393 34493: gap of 100 bp  
 \* 34493 35187: contig of 694 bp in length  
 \* 35187 35287: gap of 100 bp  
 \* 35287 35995: contig of 708 bp in length  
 \* 35995 36095: gap of 100 bp  
 \* 36095 36786: contig of 691 bp in length  
 \* 36786 36886: gap of 100 bp  
 \* 36886 37569: contig of 683 bp in length  
 \* 37569 37669: gap of 100 bp  
 \* 37669 38364: contig of 695 bp in length  
 \* 38364 38464: gap of 100 bp  
 \* 38464 39175: contig of 711 bp in length  
 \* 39175 39275: gap of 100 bp  
 \* 39275 39955: contig of 680 bp in length  
 \* 39955 40055: gap of 100 bp  
 \* 40055 40763: contig of 708 bp in length  
 \* 40763 40863: gap of 100 bp  
 \* 40863 41565: contig of 702 bp in length  
 \* 41565 42381: contig of 716 bp in length  
 \* 42381 42481: gap of 100 bp  
 \* 42481 43191: contig of 710 bp in length  
 \* 43191 43291: gap of 100 bp  
 \* 43291 44002: contig of 711 bp in length  
 \* 44002 44102: gap of 100 bp  
 \* 44102 44808: contig of 706 bp in length  
 \* 44808 44103

\* 44809 44908: gap of 100 bp  
 \* 44909 45619: contig of 711 bp in length  
 Query Match 76.8%; Score 19.2; DB 14; Length 67233;  
 Best Local Similarity 87.5%; Pred. No. 6.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 AAATTAAGAGACTATTTCGTCAA 25  
 |||||  
 Db 55694 AAATTAAGAGACTATTTCGTCAA 55671  
 |||||  
 RESULT 28  
 AC164249 125319 bp DNA linear PLN 10-AUG-2005  
 DEFINITION Medicago truncatula clone mth2-20n15, complete sequence.  
 AC164249  
 AC164249.5 GI:72095866  
 VERSION HTG.  
 KEYWORDS Medicago truncatula (barrel medic)  
 SOURCE Medicago truncatula  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 REFERENCE 1 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Medicago truncatula BAC Clone mth2-20n15  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUN-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 REFERENCE 3 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JUN-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 REFERENCE 4 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-AUG-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 COMMENT On Aug 10, 2005 this sequence version replaced gi:68342142.  
 ----- Genome Center  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----  
 FEATURES Location/Qualifiers  
 source 1. 125319  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-20n15"  
 /clone\_lib="Medicago truncatula BAC library H2"  
 ORIGIN  
 Query Match 76.8%; Score 19.2; DB 15; Length 125319;  
 Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CAAATTAAGAGACTATTTCGTGCA 24  
 |||||  
 Db 28482 CAAATTAAGATATTCATTTCGTGCA 28505

```

RESULT 29
AC149834/c
LOCUS
DEFINITION
  AC149834
  138075 bp DNA linear HTG 09-JUL-2004
  Zea mays clone ZMMBB0531A01, *** SEQUENCING IN PROGRESS ***, 7
  ordered pieces.
ACCESSION
  AC149834
VERSION
  HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEPIN.
KEYWORDS
  Zea mays
SOURCE
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 138075)
  Birren, B., Nusbaum, C., Lander, E., Butler, E., Wing, R., Bharti, A.K.
  and Messing, J.
  Zea mays, clone ZMMBB0531A01
  Unpublished
  2 (bases 1 to 138075)
  Messing, J., Abouelleil, A., Allen, N., Anderson, M., Anderson, S.,
  Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
  Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collimore, A.,
  Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S.,
  Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P.,
  Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L.,
  Grand-pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J.,
  Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
  Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
  Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P.,
  Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J.,
  Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
  Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
  O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
  Rachupka, A., Ramaeamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
  Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
  Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
  Talamas, J., Teafaye, S., Theodore, J., Topham, K., Travers, M.,
  Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
  Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
  Direct Submission
  Submitted (22-JUN-2004) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  3 (bases 1 to 138075)
  Birren, B., Nusbaum, C., Lander, E., Butler, E., Wing, R., Bharti, A.K.,
  Messing, J., Abouelleil, A., Allen, N., Anderson, M., Anderson, S.,
  Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
  Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collimore, A.,
  Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S.,
  Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P.,
  Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L.,
  Grand-pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J.,
  Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
  Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
  Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P.,
  Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J.,
  Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
  Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
  O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
  Rachupka, A., Ramaeamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
  Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
  Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
  Talamas, J., Teafaye, S., Theodore, J., Topham, K., Travers, M.,
  Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
  Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
  Direct Submission
  Submitted (09-JUL-2004) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RW/RepeatMasker.html
  
```

----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

-----  
 Bharti, AK and Messing, J: The Plant Genome Initiative at  
 Rutgers, Wakman Institute, Rutgers, The State University of New  
 Jersey, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
 (<http://pgir.rutgers.edu>)  
 Butler, E and Wing, R: Arizona Genomics Institute, Biological  
 Sciences West, 448A, P.O. Box 210086, University of Arizona,  
 Tucson, AZ 85721, USA (<http://www.genome.arizona.edu>)

----- Project Information  
 Center project name: L50346  
 Center clone name: 531\_A1

----- Consensus Information  
 This consensus is derived from a shotgun assembly that has been  
 manually curated. It is the best representation of the BAC that we  
 can generate without further laboratory work. The draft assembly  
 has been edited, and if possible, ends identified by vector as well  
 as by BAC end sequences, and contigs ordered and oriented. Bases  
 that are not Ns are either above Q20 or manually edited. This  
 assembly was performed with Arachne (Genome Res. 2002 12: 177-189;  
 Genome Res. 2003 13: 91-96). All trace files for this project are  
 available at the NCBI trace repository  
 (<http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi?>). An exact list of  
 reads used in this assembly are available at  
<http://www.broad.mit.edu/annotation/plants/maize/randomclones.html>.

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 1 28159: contig of 28159 bp in length  
 \* 28160 28259: gap of 100 bp  
 \* 28260 42651: contig of 14392 bp in length  
 \* 42652 42751: gap of 100 bp  
 \* 42752 50861: contig of 8110 bp in length  
 \* 50862 50961: gap of 100 bp  
 \* 50962 59561: contig of 8600 bp in length  
 \* 59562 104318: contig of 44657 bp in length  
 \* 104319 104418: gap of 100 bp  
 \* 104419 113136: contig of 8718 bp in length  
 \* 113137 113236: gap of 100 bp  
 \* 113237 138075: contig of 24839 bp in length.

#### FEATURES

Location/Qualifiers  
 1..138075  
 /organism="Zea mays"  
 /molecule="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBB0531A01"  
 /clone\_lib="CUGI Maize B73 BAC HindIII Library"  
 28160..28259  
 /estimated\_length=100  
 42652..42751  
 /estimated\_length=100  
 50862..50961  
 /estimated\_length=100  
 59562..59561  
 /estimated\_length=100  
 104319..104418  
 /estimated\_length=100  
 113137..113236  
 /estimated\_length=100

ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 138075;  
 Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 14350 CAAATTAAGAGACTATTCGTGCA 14327

RESULT 30  
 AC073928 138943 bp DNA linear PRI 30-APR-2005  
 LOCUS  
 DEFINITION Homo sapiens BAC clone RP11-745P9 from 2, complete sequence.  
 AC073928  
 ACCESSION  
 VERSION AC073928.5 GI:16950374  
 HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 AUTHORS Kozlowicz, A., Spalding, L. and Cedroni, M.  
 TITLE The sequence of Homo sapiens BAC clone RP11-745P9  
 JOURNAL Unpublished (2001)  
 REFERENCE  
 AUTHORS 2 (bases 1 to 138943)  
 TITLE Direct Submission  
 JOURNAL Waterston, R.H.  
 REFERENCE  
 AUTHORS Submitted (05-JUL-2000) Genome Sequencing Center, Washington  
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 JOURNAL MO 63108, USA  
 REFERENCE  
 AUTHORS 3 (bases 1 to 138943)  
 TITLE Direct Submission  
 JOURNAL Waterston, R.H.  
 REFERENCE  
 AUTHORS Submitted (16-NOV-2001) Genome Sequencing Center, Washington  
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 JOURNAL MO 63108, USA  
 REFERENCE  
 AUTHORS 4 (bases 1 to 138943)  
 TITLE Direct Submission  
 JOURNAL Waterston, R.H.  
 REFERENCE  
 AUTHORS Submitted (03-JAN-2002) Genome Sequencing Center, Washington  
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 JOURNAL MO 63108, USA  
 REFERENCE  
 AUTHORS 5 (bases 1 to 138943)  
 TITLE Direct Submission  
 JOURNAL Waterston, R.  
 REFERENCE  
 AUTHORS Submitted (09-JAN-2002) Department of Genetics, Washington  
 TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 JOURNAL 6 (bases 1 to 138943)  
 REFERENCE  
 AUTHORS Wilson, R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-APR-2005) Genome Sequencing Center, Washington  
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 JOURNAL MO 63108, USA  
 COMMENT On Nov 16, 2001 this sequence version replaced gi:16756379.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0745P09  
 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. Wes Warren,  
 Department of Genetics, Washington University, St. Louis MO. For  
 additional information about the map position of this sequence, see  
 http://genome.wustl.edu

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male  
 donor, as described by Oosagawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Tateno, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
 and coworkers at http://www.chori.org  
 VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-326J14, 2000 bp overlap.  
 Actual start of this clone is at base position 1 of RP11-745P9;  
 Actual end is at base position 48928 of RP11-326J14.

Single plasmid regions exist between 6749 and 67933; 82507 and  
 82581.

## FEATURES

Location/Qualifiers

1..138943

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="2"

/clone="RP11-745P9"

/clone\_lib="RPCI-11"

10825..11105

/note="CpG island (%GC=61.9, o/e=0.79, #CpGs=22)"

75901..76274

/note="CpG island (%GC=74.3, o/e=0.54, #CpGs=33)"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 138943;  
 Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCA 25  
 ||||| ||||| ||||| ||||| |||||

Db 3560 AAATTAAGAGACTATTCGTGCA 3583

## RESULT 31

AC012362/c

LOCUS

DEFINITION Homo sapiens BAC clone RP11-404D15 from 2, complete sequence.

AC012362

ACCESSION

VERSION AC012362.7 GI:16306515

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homnidae; Homo.

REFERENCE 1 (bases 1 to 156618)

AUTHORS Cotton, M., Maupin, R., Phillips, A., Elliott, G., Boyer, E. and Paulson, E.

TITLE The sequence of Homo sapiens BAC clone RP11-404D15

JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 156618)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-1999) Genome Sequencing Center, Washington

REFERENCE 3 (bases 1 to 156618)

AUTHORS MO 63108, USA

**TITLE**  
JOURNAL  
Submitted (22-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
**REFERENCE**  
AUTHORS  
4 (bases 1 to 156618)  
Waterston,R.H.  
**TITLE**  
JOURNAL  
Submitted (23-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
**REFERENCE**  
AUTHORS  
5 (bases 1 to 156618)  
Waterston,R.  
**TITLE**  
JOURNAL  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
**REFERENCE**  
AUTHORS  
6 (bases 1 to 156618)  
Wilson,R.K.  
**TITLE**  
JOURNAL  
Submitted (21-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
**COMMENT**  
On Oct 22, 2001 this sequence version replaced gi:15029477.  
-----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
-----  
Summary Statistics  
-----  
Center project name: H\_NH0404D15  
-----

**NOTICE:**

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

**MAPPING INFORMATION:**

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

**SOURCE INFORMATION:**

The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Geogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,  
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

**NEIGHBORING SEQUENCE INFORMATION:**

The clone sequenced to the left is RP11-140C4, 2000 bp overlap; the  
clone sequenced to the right is RP11-299P7, 2000 bp overlap.  
Actual start of this clone is at base position 105844 of  
RP11-140C4; actual end is at base position 22547 of RP11-299P7.

Polymorphisms have been identified between AC019185 and AC012362.

**FEATURES**  
Source

1..156618  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-404D15"  
/clone\_lib="RPCI-11"  
615..1993

gene

**misc\_feature**  
56693..57028  
/note="CpG island (GC=71.7, o/e=0.64, #CpGs=29)"  
**misc\_feature**  
73890..74689  
/note="CpG island (GC=70.0, o/e=0.66, #CpGs=59)"  
**misc\_feature**  
103058..103707  
/note="CpG island (GC=69.7, o/e=0.81, #CpGs=70)"  
**gene**  
103347..156406  
/gene="PTHR2"  
**mRNA**  
Join(103347..103563,124646..124748,133982..134092,  
134205..134326,138809..138906,139793..139982,  
141179..141332,147194..147254,156340..156406)  
/gene="PTHR2"  
**CDS**  
Join(103489..103563,124646..124748,133982..134092,  
134205..134326,138809..138906,139793..139982,  
141179..141332,147194..147254,156340..156406)  
/gene="PTHR2"  
/note="Homo sapiens parathyroid hormone receptor 2  
(PTHR2), mRNA.; H\_NH0404D15.2  
This gene was based on gi(4826953)  
Continues as H\_NH0299P07.1"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AA14871.1"  
/db\_xref="GI:62822322"  
/translation="MAGLGASLVHVGWMLGSCLLARQLDSGDTTIEQIVLVKHA  
KVQCELNITAOQECCGNCPEWDGLICWPRGTGKISAVPCPIYDFNKHGVAFRH  
KNPGLTDFHSLNKTNYVHMLFVSPMLRATISFVKDRVVAHIGVKELSLIMQDD  
VAIIIGVFRHLCTRNVIHMLFVSPMLRATISFVKDRVVAHIGVKELSLIMQDD  
PQNSIEATSDKSYIGCKIAVMFVFLATNYVILVEGLVHLNLIFFVAFSDTKYL  
WGFILIGWFFPAFVAWAVARATLADARCWELSGDKIKWYIQAPILAIG"

**ORIGIN**

Query Match 76.8%; Score 19.2; DB 8; Length 156618;  
Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 2 AAATTAAGAGACTATTCTGTCAA 25  
|||||  
Db 97208 AAATTAAGAGACTATTCTGTCAA 97185

**RESULT 32**  
AC009306  
LOCUS  
DEFINITION Homo sapiens BAC clone RP11-290K24 from 2, complete sequence.  
ACCESSION AC009306  
VERSION AC009306.5 GI:10716644  
KEYWORDS HTG.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
1 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (13-AUG-1999) Genome Sequencing Center, Washington

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
2 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (26-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (27-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (28-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (29-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (30-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
7 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (31-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
8 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (05-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
9 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (07-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
10 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (12-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
11 (bases 1 to 176602)  
Waterston, R.  
Direct Submission  
Submitted (07-OCT-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
12 (bases 1 to 176602)  
Waterston, R.H.  
Direct Submission  
Submitted (25-MAR-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
13 (bases 1 to 176602)  
Waterston, R.  
Direct Submission  
Submitted (29-OCT-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
14 (bases 1 to 176602)

AUTHORS  
TITLE  
JOURNAL

Wilson, R.K.  
Direct Submission  
Submitted (05-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Oct 7, 2000 this sequence version replaced gi:8099093.  
FEATURES  
source  
1. 176602  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-290K24"  
/clone\_lib="RPC1-11"  
142050..142278  
misc\_feature  
/note="CpG island (GC=60.3, o/e=0.91, #CpGs=20)"  
146574..146859  
misc\_feature  
/note="CpG island (GC=61.2, o/e=0.70, #CpGs=21)"  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 176602;  
Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 96881 AATTAAGAGACTATTCGTGCTA 96904  
|||||  
RESULT 33  
AC154981  
LOCUS  
DEFINITION  
AC154981  
AC154981.2 GI:68265074  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Bos taurus (cow)  
SOURCE  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 182536)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angilano, D.,  
Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, C., Deramo, S., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Loulseghe, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Muniadase, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., NewCon, N., Nguyen, N., Norris, S., Parks, K., Nwackelme, O., Okwunonu, G., Olarnpunaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, T., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trajcs, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

# TITLE JOURNAL REFERENCE AUTHORS

Direct Submission  
Unpublished  
2 (bases 1 to 182536)

Worley, K.C.

# TITLE JOURNAL

Submitted (06-JAN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 182536)

Cow Genome Sequencing Consortium.

# TITLE JOURNAL

Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# COMMENT

On Jun 28, 2005 this sequence version replaced gi:57164456. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FBZ2  
Center clone name: CH240-27D11  
----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 178361 bases at least Q40

Consensus quality: 179461 bases at least Q30

Consensus quality: 180552 bases at least Q20

Estimated insert size: 182232; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 68663: contig of 68663 bp in length  
\* 68664 68713: gap of 50 bp  
\* 68714 105062: contig of 36349 bp in length  
\* 105063 105555: gap of 493 bp  
\* 105556 137308: contig of 31753 bp in length  
\* 137309 137358: gap of 50 bp  
\* 137359 140589: contig of 3231 bp in length  
\* 140590 140639: gap of 50 bp  
\* 140640 14837: contig of 4198 bp in length  
\* 14838 14887: gap of 50 bp  
\* 14888 147571: contig of 2684 bp in length  
\* 147572 147621: gap of 50 bp  
\* 147622 169515: contig of 21894 bp in length  
\* 169516 169565: gap of 50 bp  
\* 169566 173031: contig of 3466 bp in length  
\* 173032 173131: gap of unknown length  
\* 173132 174955: contig of 1824 bp in length  
\* 174956 175005: gap of 50 bp  
\* 175006 179124: contig of 4119 bp in length  
\* 179125 179224: gap of unknown length  
\* 179225 180668: contig of 1444 bp in length  
\* 180669 180768: gap of unknown length  
\* 180769 182536: contig of 1768 bp in length.  
\* 180770

# FEATURES

Location/Qualifiers  
Source

1. 182536  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-27D11"  
68664..68713  
/estimated\_length=50  
105063..105555  
/estimated\_length=493  
137309..137358  
/estimated\_length=50  
140590..140639  
/estimated\_length=50  
144838..144887  
/estimated\_length=50  
147572..147621  
/estimated\_length=50  
169516..169565  
/estimated\_length=50  
173032..173131  
/estimated\_length=unknown  
174956..175005  
/estimated\_length=50  
179125..179224  
/estimated\_length=unknown  
180669..180768  
/estimated\_length=unknown

# ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 182536;  
Best Local Similarity 87.5%; Pred. No. 5,3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGCAA 25  
|||||

Db 25615 AAAATGAAGAGACTATTCTGCAA 25638  
|||||

# RESULT 34

AC146141

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC146141 182547 bp DNA linear PRI 29-OCT-2003  
Pan troglodytes BAC clone RP43-3A23 from 7, complete sequence.

AC146141

AC146141.2 GI:35073670

HTG.

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;





80 - 89 : 62232  
90 - 99 : 79090  
-----  
Percentage of bases with a quality value >= 40 : 99 %.

## FEATURES

source

Location/Qualifiers  
1..183491  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-561B11"  
/clone\_lib="RPCI-11"  
17109..17210  
/note="matching EMBL:X59417  
RHdb:RH79927  
dbSTS:STS56350  
Identified using the e-PCR software (G. Schuler)"  
STS  
101390..101659  
/note="matching EMBL:M69043  
RHdb:RH69167  
dbSTS:STS49057  
Identified using the e-PCR software (G. Schuler)"  
STS  
101471..101667  
/note="matching EMBL:G13563  
RHdb:RH7865  
dbSTS:STS13729  
Identified using the e-PCR software (G. Schuler)"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 183491;  
Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAAGAGACTATTCTGCA 24

Db 146159 CAATTAAAGAGACTTTTCATCA 146136

## RESULT 36

AC157660

LOCUS

AC157660 192026 bp DNA linear ROD 23-APR-2005  
Mus musculus BAC clone RP23-263116 from chromosome 1, complete  
sequence.

## DEFINITION

AC157660

VERSION

AC157660.2 GI:60543592

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 192026)

REFERENCE

Hlatka M. and Kozlowski A.

The sequence of Mus musculus BAC clone RP23-263116

Unpublished (2001)

2 (bases 1 to 192026)

REFERENCE

Wilson R.K.

Direct Submission

Submitted (24-FEB-2005) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

3 (bases 1 to 192026)

REFERENCE

Wilson R.K.

Direct Submission

Submitted (05-MAR-2005) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 192026)

REFERENCE

Wilson R.K.

Direct Submission

Submitted (23-APR-2005) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

On Mar 5, 2005 this sequence version replaced gi:60223388.  
----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

----- Summary Statistics

----- Center project name: M\_BA0263116

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.

If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The BAC Library has been constructed by Kazutoyo Oseawa and Minako Tateo in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES  
source

Location/Qualifiers  
1..192026  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="RP23-263116"  
/clone\_lib="RPCI-23"

misc\_feature

1..13  
/note="Sequence derived from PCR product of project DNA"

misc\_feature

282..434  
/note="Sequence derived from PCR product of project DNA"

misc\_feature

21398..21399  
/note="Bacterial transposon insertion in clone excised here"

unsure

123614..123676  
/note="Sequence derived from one plasmid subclone."

unsure

158282..158374  
/note="Unresolved simple sequence repeat."

unsure

171648..171695  
/note="Sequence derived from one plasmid subclone."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 192026;  
Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTCTGCA 25

|||||

```
Db 102489 AATAAAGGAGACTATTAGTGCA 102512

RESULT 37
CNS01RGW/c 194173 bp DNA linear PRI 04-MAY-2001
LOCUS Human chromosome 14 DNA sequence BAC R-355C3 of library RPI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL160231
VERSION AL160231.4 GI:13992187
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
REFERENCE 1 (bases 1 to 194173)
AUTHORS Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissbach, J.
TITLE Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194173)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT On May 8, 2001 this sequence version replaced gi:13016588.
----- Genom Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-561B11
Downstream BAC (overlapping the SP6 end) : R-102C24 (AC=AL137818)
Assembly program: Phrap; version 2.0
Quality coverage: 7.06x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 :
30 - 39 : 9
40 - 49 : 115
50 - 59 : 2501
60 - 69 : 7708
70 - 79 : 8382
80 - 89 : 18066
90 - 99 : 49946
107446
-----
Percentage of bases with a quality value >= 40 : 99 %.
-----
FEATURES
source
1. 194173
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="R-355C3"
/clone_lib="RPI-11"
2179..72308
/note="matching EMBL:G07705
RHdb:RH32619
dbSTS:STS32292
Identified using the e-PCR software (G. Schuler)"
95324..95487
STS
```

/note="matching EMBL:Z39869

RHdb:RH53631

dbSTS:STS17739

Identified using the e-PCR software (G. Schuler)"

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 194173;  
Best Local Similarity 87.5%; Pred. No. 5.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGTGCA 24

Db 3237 CAATTAAGAGACTTTTCATTC 3214

#### RESULT 38

AC136716

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 259603)

Birren, B., Nusbaum, C. and Lander, E.

Unpublished

REFERENCE 2 (bases 1 to 259603)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genom Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information



Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCTGCA 24  
||||| ||||||| ||||||| |||

Db 150487 CAAATCAAGAGATTATTCGAGCA 150510

RESULT 39  
AC152434/c  
LOCUS AC152434.3 309502 bp DNA linear HTG 01-JUL-2005  
DEFINITION Bos taurus clone CH240-5P17, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 59  
unordered pieces.

AC152434.3 GI:68227043  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Bos taurus (cow)

SOURCE Bos taurus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 309502)  
Muzny, D. Marie., Metzger, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, P., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hult, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankerville, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwakoleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,  
Pruzzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D.,  
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Weinstock, G., and Gibbs, R. A.

Direct Submission  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

2 (bases 1 to 309502)  
Worley, K.C.  
Direct Submission  
Submitted (08-NOV-2004) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 309502)  
Cow Genome Sequencing Consortium.  
Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 26, 2005 this sequence version replaced gi:58037939.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: FAXM  
Center clone name: CH240-5P17  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 285635 bases at least Q40  
Consensus quality: 290153 bases at least Q30  
Consensus quality: 294335 bases at least Q20  
Estimated insert size: 285852; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 59 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
\* 11017: contig of 11017 bp in length  
\* 11629: gap of 612 bp  
\* 11630: contig of 3854 bp in length  
\* 15484: gap of 181 bp  
\* 15484: contig of 9627 bp in length  
\* 25291: gap of 50 bp  
\* 25292: contig of 3855 bp in length  
\* 25342: gap of 50 bp  
\* 29197: contig of 1945 bp in length  
\* 29247: gap of 50 bp  
\* 31192: contig of 8802 bp in length  
\* 31242: gap of 50 bp  
\* 40043: contig of 6544 bp in length  
\* 40094: gap of 50 bp  
\* 46638: contig of 11709 bp in length  
\* 46688: gap of 50 bp  
\* 58397: contig of 4732 bp in length  
\* 58447: gap of 50 bp  
\* 63179: contig of 7463 bp in length  
\* 63229: gap of 50 bp  
\* 70691: contig of unknown length  
\* 70692: gap of 4004 bp in length  
\* 70795: contig of 4004 bp in length

```
* 74796 74845: gap of 50 bp
* 74846 80620: contig of 5775 bp in length
* 80621 80670: gap of 50 bp
* 80671 86253: contig of 5583 bp in length
* 86254 86303: gap of 50 bp
* 86304 86588: contig of 10355 bp in length
* 86589 96758: gap of unknown length
* 96759 97822: contig of 1064 bp in length
* 97823 97921: gap of 99 bp
* 97922 104351: contig of 6430 bp in length
* 104352 104401: gap of 50 bp
* 104402 115482: contig of 11081 bp in length
* 115483 115532: gap of 50 bp
* 115533 128633: contig of 13101 bp in length
* 128634 128733: gap of unknown length
* 128734 133221: contig of 4488 bp in length
* 133222 133271: gap of 50 bp
* 133272 139579: contig of 6308 bp in length
* 139580 139629: gap of 50 bp
* 139630 148562: contig of 8933 bp in length
* 148563 148805: gap of 243 bp
* 148806 161450: contig of 12645 bp in length
* 161451 161500: gap of 50 bp
* 161501 165216: contig of 3716 bp in length
* 165217 165266: gap of 50 bp
* 165267 167490: contig of 2224 bp in length
* 167491 168006: gap of 516 bp
* 168007 172330: contig of 4324 bp in length
* 172331 173367: gap of 1037 bp
* 173368 176298: contig of 2931 bp in length
* 176299 176892: gap of 594 bp
* 176893 184384: contig of 7492 bp in length
* 184385 184434: gap of 50 bp
* 184435 192053: contig of 7619 bp in length
* 192054 192103: gap of 50 bp
* 192104 193660: contig of 1557 bp in length
* 193661 195706: gap of 2046 bp
* 195707 202230: contig of 6524 bp in length
* 202231 202280: gap of 50 bp
* 202281 211945: contig of 9665 bp in length
* 211946 211995: gap of 50 bp
* 211996 232014: contig of 20019 bp in length
* 232015 232064: gap of 50 bp
* 232065 234686: contig of 2622 bp in length
* 234687 234786: gap of unknown length
* 234787 237743: contig of 2957 bp in length
* 237744 238139: gap of 396 bp
* 238140 239574: contig of 1435 bp in length
* 239575 239624: gap of 50 bp
* 239625 246849: contig of 7225 bp in length
* 246850 246899: gap of 50 bp
* 246900 250034: contig of 3135 bp in length
* 250035 250134: gap of unknown length
* 250135 251145: contig of 1011 bp in length
* 251146 252472: gap of unknown length
* 252473 252572: gap of unknown length
* 252573 254244: contig of 1672 bp in length
* 254245 254344: gap of unknown length
* 254345 255355: contig of 1011 bp in length
* 255356 255455: gap of unknown length
* 255456 256842: contig of 1387 bp in length
* 256843 256942: gap of unknown length
* 256943 258008: contig of 1066 bp in length
```

Query Match 76.8%; Score 19.2; DB 14; Length 309502;  
Best Local Similarity 87.5%; Pred. No. 4.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTCTGCA 24  
|||||  
DB 34653 CAAATGAAGAGAGACTATTAGTGA 34630

RESULT 40  
BX901880\_2/c  
WPCOMMENT

Sequence split into 4 fragments LOCUS BX901880 Accession BX901880

Fragment Name	Begin	End
BX901880_0	1	110000
BX901880_1	100001	210000
BX901880_2	200001	310000
BX901880_3	300001	370132

Continuation (3 of 4) of BX901880 from base 200001 (BX901880 Danio rerio clone DKBY-109,

Query Match 75.2%; Score 18.8; DB 14; Length 110000;  
Best Local Similarity 90.9%; Pred. No. 8.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGAGACTATTCTGCA 25  
|||||  
DB 91179 ATTAAGAGAGACTATTCTGCA 91158

RESULT 41

BX323460/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfiah-help@sanger.ac.uk](mailto:zfiah-help@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Zebrafish PUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where

this is found the longest good quality representation will be

submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhirong Bao and Sean Eddy, submitted), and those

beginning 'drr' were identified by Rick Waterman (Stephen Johnson

submitted).

lab, WashU). For further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEYP-47G11  
 is from a Zebrafish BAC library  
 VECTOR: pIndigoBAC-5.

FEATURES  
 source  
 Location/Qualifiers  
 1..202017  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEYP-47G11"  
 /clone\_lib="DanioKeyypilot"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 202017;  
 Best Local Similarity 90.9%; Pred. No. 7.6e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTGTCGCA 24  
 ||||| ||||| ||||| |||||  
 Db 19306 AATTAAGAGACTATTGTCGCA 19285

## RESULT 42

CR391929/c  
 LOCUS  
 DEFINITION  
 Zebrafish DNA sequence from clone DKEYP-86B5, complete sequence.

ACCESSION  
 CR391929  
 VERSION  
 CR391929.9 GI:60219381

KEYWORDS  
 HTG.

SOURCE  
 Danio rerio (zebrafish)

## ORGANISM

Danio rerio  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.

## REFERENCE

1 (bases 1 to 223724)

## AUTHORS

Glithero, R.

## JOURNAL

Submitted (22-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

On Feb 23, 2005 this sequence version replaced gi:59891260.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: zfish-help@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived  
 Zebrafish pUC subclones occasionally display inconsistency over the  
 length of mononucleotide A/T runs and conserved TA repeats. Where  
 this is found the longest good quality representation will be  
 submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat  
 discovery system (Zhifeng Bao and Sean Eddy, submitted), and those  
 beginning 'drr' were identified by Rick Waterman (Stephen Johnson

lab, WashU). For further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEYP-86B5  
 is from a Zebrafish BAC library  
 VECTOR: pIndigoBAC-5.

FEATURES  
 source  
 Location/Qualifiers  
 1..223724  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEYP-86B5"  
 /clone\_lib="DanioKeyypilot"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 223724;  
 Best Local Similarity 90.9%; Pred. No. 7.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTGTCGCAA 25  
 ||||| ||||| ||||| |||||  
 Db 84895 ATTAAGAGACTATTGTCGCAA 84874

## RESULT 43

AC098905/c  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-163J19, WORKING DRAFT SEQUENCE, 2  
 unordered pieces.

ACCESSION  
 AC098905  
 VERSION  
 AC098905.6 GI:30521215

KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE  
 Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.

## REFERENCE

1 (bases 1 to 225849)

## AUTHORS

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Bialwalo, K., Blair, J., Blankenburg, K., Blyth, K., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Caldarone, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, Z., Chu, J.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falle, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guavara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhsu, L., Loulsegue, H., Lozado, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: GIOC  
Center clone name: CH230-163J19  
----- Summary Statistics -----  
Consensus quality: 212145 bases at least Q40  
Consensus quality: 214496 bases at least Q30  
Consensus quality: 216260 bases at least Q20  
Estimated insert size: 223073; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

1 222215: contig of 222215 bp in length  
222216 222315: gap of unknown length  
222316 225849: contig of 3534 bp in length.  
Location/Qualifiers  
1 .225849  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

FEATURES  
source

/clone="CH230-163J19"  
1 .1302  
/note="wgs\_contig"  
214638 .217165  
/note="wgs\_contig"  
220987 .222215  
/note="wgs\_contig"  
222216 .222315  
/estimated\_length=unknown

ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 225849;  
Best Local Similarity 90.9%; Pred. No. 7.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTCTGTC 23  
|||||  
Db 173065 AAATTAAGAGACTATTCTGTC 173044

RESULT 44

BX682552/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zf18h-help@sanger.ac.uk](mailto:zf18h-help@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e. phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/c\\_elegans/wormpep](http://www.sanger.ac.uk/projects/c_elegans/wormpep)

Cloned pUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where

this is found the longest good quality representation will be

submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhirong Bao and Sean Eddy, submitted), and those

beginning 'drr' were identified by Rick Waterman (Stephen Johnson

lab, Washu). For further information see



http://www.sanger.ac.uk/Projects/D\_rerio/fishmask.shtml DKEY-120C6  
is from a Zebrafish BAC library

VECTOR: pindigoBAC-5

Location/Qualifiers

1. 228044  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-120C6"  
/clone\_lib="DanioKey"

#### FEATURES

source

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 228044;

Best Local Similarity 90.9%; Pred. No. 7.4e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCTGTCAA 25

Db 44750 ATTAAGAAAGTATTCTGTCAA 44729

#### RESULT 45

AC120660/c

LOCUS

AC120660 231498 bp DNA linear HTG 20-NOV-2002

Rattus norvegicus clone CH230-33A2, WORKING DRAFT SEQUENCE, 2

unordered pieces.

#### ACCESSION

AC120660.6 GI:25137880

HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

#### SOURCE

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Murioidea; Muridae; Murinae; Rattus.

1 (bases 1 to 231498)

Muzny, D. Marie., Metzker, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandarainke, D., Barber, M., Barnstead, M., Benahmed, P.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,

Hollins, B., Howells, S., Hulvyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensuewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J.,

Mareshwari, M., Mahindarte, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokemele, O., Okwunu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plommer, F., Poidexter, A., Popovic, D., Primus, E., Pu, L.-L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

#### COMMENT

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 231498)

Worley, K.C.

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231498)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23269608.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GXLB

Center clone name: CH230-33A2

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 224823 bases at least Q40

Consensus quality: 225934 bases at least Q30

Consensus quality: 226530 bases at least Q20

Estimated insert size: 229373; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 229826: contig of 229826 bp in length

\* 229827 229926: gap of unknown length

\* 229927 231498: contig of 1572 bp in length.

Location/Qualifiers

1..231498

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-33A2"

#### FEATURES

source

score  $\geq 20$ , at least 10% of its base calls must satisfy SNQS(30,25)(single strand NQS, the base in question has Phred score  $\geq 30$ , the surrounding 10 bases in the read have Phred score  $\geq 25$ ), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate  $> 0.01$  were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments ( $\geq 95\%$  bases of read A and  $\geq 95\%$  bases of read B were placed at the same locus of human genome) were discarded.

```

FEATURES
source

STS
ORIGIN

```

Query Match	74.4%	Score 18.6;	DB 10;	Length 575;
Best Local Similarity	84.0%;	Pred. No. 3e+03;		
Matches	21;	Conservative	0; Mismatches	4; Indels
Gaps	0;			
Qy	1	CRAATTTAAAGAGACTATTCTGTCAA	25	
Db	419	CRAATTTAAAGAGATTATTTTTGCAA	395	

RESULT	47
BV552932/c	
LOCUS	694 bp DNA linear STS 09-APR-2005
DEFINITION	S21P694F0BP3.T0 Noemie Pan troglodytes troglodytes STS genomic, sequence tagged site.
ACCESSION	BV552932
VERSION	BV552932.1 GI:62443952
KEYWORDS	STS.
SOURCE	Pan troglodytes troglodytes
ORGANISM	Pan troglodytes troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Pan. 1 (bases 1 to 694)
REFERENCE	Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and Jaffe,D.B.
AUTHORS	Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome
TITLE	Unpublished (2005)
JOURNAL	
COMMENT	

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 694  
Protocol:  
23, 021, 928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps

```

misc_feature      1. .1414
/note="wgs_end_extension
clone_end:Sp6"
3437. .4234
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BH351289"
complement(224954. .225701)
/note="Clone_boundary
clone_end:T7
site:
end_sequence:BH351287"
227043. .229826
/note="wgs_end_extension
clone_end:T7"
229827. .229926
/estimated_length=unknown

ORIGIN
Query Match      75.2%; Score 18.8; DB 14; Length 231498;
Best Local Similarity 90.9%; Pred. No. 7,4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Y      2 AAATTAAAGAGACTATTCTGTC 23
      ||| ||||| ||||| |||
      87242 AAATAAAGAGACTATTCTGTC 87221

```

```

Query Match      75.2%; Score 18.8; DB 14; Length 231498;
Best Local Similarity 90.5%; P: 7.4e-02;
Matches 20; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 46	BV515800	575 bp	DNA	linear	STS 08-APR-2005
OCUS	BV515800/c	rfp75e02.gi	Clint Pan troglodytes	versus STS genomic,	sequence
DEFINITION			tagged site.		
ACCESSION	BV515800				
VERSION	BV515800.1	GI:62393558			
KEYWORDS	STS.				
SOURCE	Pan troglodytes	versus			
ORGANISM	Pan troglodytes	versus			
	Eukaryota;	Metazoa;	Chordata;	Cranialia;	Vertebrata; Euteleostomi;
	Mammalia;	Eutheria;	Euarchontoglires;	Primates;	Catarrhini;
	Hominidae;	Pan.			
REFERENCE	1	(bases 1 to 575)			
AUTHORS	Mikkelson,T.S.,	Hillier,W.L.,	Eichler,E.E.,	Zody,M.C.	and
	Jaffe,D.B.				
TITLE	Initial Sequence of the Chimpanzee Genome and Comparison with the				
	Human Genome				
JOURNAL	Unpublished (2005)				

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 575  
Protocol:  
23, 021, 928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have phred

(Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score  $\geq 20$ , at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score  $\geq 30$ , the surrounding 10 bases in the read have Phred score  $\geq 25$ ), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate  $> 0.01$  were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments ( $> 95\%$  bases of read A and  $> 95\%$  bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
1. .694  
Location/Qualifiers  
/organism="Pan troglodytes troglodytes"  
/mol\_type="genomic DNA"  
/sub\_species="troglodytes"  
/db\_xref="taxon:37011"  
/clone\_lib="Noemie"  
<1. .>694

STS  
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 694;  
Best Local Similarity 84.0%; Pred. No. 2.9e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
|||||  
Db 609 CAAATTAAAGAGATTATTAGCAA 585

RESULT 48  
BV529432/c  
LOCUS 703 bp DNA linear STS 08-APR-2005  
DEFINITION G591P613281P66.T0 Clint Pan troglodytes verus STS genomic, sequence tagged site.  
ACCESSION BV529432  
VERSION BV529432.1 GI:62407216  
KEYWORDS STS.  
SOURCE Pan troglodytes verus  
ORGANISM Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.

REFERENCE 1 (bases 1 to 703)  
AUTHORS Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and Jaffe, D.B.  
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome  
JOURNAL Unpublished (2005)  
COMMENT Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu

Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 703

Protocol:  
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score  $\geq 20$ , at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score  $\geq 30$ , the surrounding 10 bases in the read have Phred score  $\geq 25$ ), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate  $> 0.01$  were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments ( $> 95\%$  bases of read A and  $> 95\%$  bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
1. .703  
Location/Qualifiers  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1. .>703

STS  
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 703;  
Best Local Similarity 84.0%; Pred. No. 2.9e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
|||||  
Db 540 CAAATTAAAGAGATTATTAGCAA 516

RESULT 49  
BV521567/c  
LOCUS 940 bp DNA linear STS 08-APR-2005  
DEFINITION G591P66182RD10.T0 Clint Pan troglodytes verus STS genomic, sequence tagged site.

ACCESSION BV521567  
VERSION BV521567.1 GI:62399337  
KEYWORDS STS.  
SOURCE Pan troglodytes verus  
ORGANISM Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.

REFERENCE 1 (bases 1 to 940)  
AUTHORS Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and Jaffe, D.B.  
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the

JOURNAL COMMENT	Human Genome Unpublished (2005)		AV513464.1 GI:47119012	
	Contact: Michael C. Zody Broad Institute of MIT and Harvard 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172580933 Fax: 6172580903 Email: mczody@broad.mit.edu Primer A: No sequence submitted Primer B: No sequence submitted STS size: 940 Protocol: 23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25)(single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.		mitochondrion Stalotropa fairmairii Stalotropa fairmairii Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Membracidae; Membracidae; Stalotropa. 1 (bases 1 to 1812) Lin, C.-P., Danforth, B.N. and Wood, T.K. Molecular phylogenetics and evolution of maternal care in membracine treehoppers Syst. Biol. 53 (3) 400-421 (2004) 2 (bases 1 to 1812) Lin, C.-P., Danforth, B.N. and Wood, T.K. Direct Submission Submitted (29-DEC-2003) Entomology, Cornell University, Comstock Hall, Ithaca, NY 14853-0901, USA Location/Qualifiers 1..1812 /organism="Stalotropa fairmairii" /organelle="mitochondrion" /mol_type="genomic DNA" /isolate="STAL" /db_xref="taxon:269982" <1..1232 /gene="COI" /gene="COI" /note="TAA stop codon is completed by the addition of 3' A residues to the mRNA" /codon_start=2 /transl_except=(pos:1232,aa:TERM) /transl_table=5 /product="cytochrome oxidase subunit I" /protein_id="AAT11466.1" /db_xref="GI:47119013" /translation="FLPSPILLLTSTAGGAGTGTVPPLSSNVAHSGPSVDLVI PSLHLAGASSITGLAINLITIMNRCESKMLDQSLFTWIFITAFELLILSLPVLAG ITMLGFRNFSTFFDSGGDPIYLQHLFWFGHPVYILILPFGVISHIVQESG KNSFGYIGMMYAMISIGGLGFIWAHMFVTGMDVDTRATSTWIIAIPGKIF SWLATMGSCILKLSVIMWAGSFLFTVGLTGLIVLANSIDILHDITYVVAHFY VLSMGAVFTIMAGHIWYPLITGTVMNNKMLKIQFFSMFIGVNNTPFPQHLGNGMP RRYSDYDPSYMLNWNVSSFGSVLSISITFMIFVIESLMSKRLSLFSFMNSSIEM QEPKHSYSELPLVN" 1233..1298 /product="tRNA-Leu" 1299..>1812 /gene="COII" 1299..>1812 /gene="COII" /gene="COII" /codon_start=1 /transl_table=5 /product="cytochrome oxidase subunit II" /protein_id="AAT11467.1" /db_xref="GI:47119014" /translation="MANIKMSFODPSSNNMQLVFFNDHVMIMILITMVLVYLMFSS SMVSLNRLNLEQGLFIWTLPAILLVFIALPSKILYLLETISPMITKTLGH QWYSYEFSDFTKIEFDSYKPMNDLNQQBEKFLVDNRLVVPYTPIRILASSLDVV HSWTIPSLGVR"	
FEATURES	Source		Query Match	
	1..940 /organism="Pan troglodytes verus" /mol_type="genomic DNA" /sub_species="verus" /db_xref="taxon:37012" /clone_lib="Clint" <1..940		Best Local Similarity 74.4%; Score 18.6; DB 2; Length 1812; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
STS	QY		1 CAATTAAGAGACTATTCTGCGAA 25	
	Db		665 CAATTAAGAGACTATTCTGCGAA 641	
ORIGIN	Query Match		74.4%; Score 18.6; DB 2; Length 1812;	
	Best Local Similarity 84.0%; Pred. No. 2.7e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		Best Local Similarity 84.0%; Pred. No. 2.4e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
LOCUS	QY		1 CAATTAAGAGACTATTCTGCGAA 25	
	Db		1083 CTAATTAAGAGACTACTCTTCAA 1059	
ACCESSION	RESULT 50		AY064171	
	AY513464/c		AY064171	
DEFINITION	Stalotropa fairmairii isolate STAL cytochrome oxidase subunit I (COI) gene, partial cds; tRNA-leu gene, complete sequence; and cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial.		2461 bp DNA linear BCT 26-DEC-2001	
	AY513464		AY064171	



```

VAHDLTPSDTAQLNKKFKVAFVTDIGRTSHSAIMARTLEIPAVLTGTTNNITELVKDGD
ILAVSGITGEVVIINPTESQIAIEFAAGGEDYAKQAEWAQLKADAPTVTADGKHFELAA
IGTPKQGVNDNGAEAVGLYRTFLYMSQDFTTEQDYEAAYKAVLEGMNGKPVVVR
TMDIGGDKELPYFDLPKEMNPFGLGYRALRISISETGNQMFRTQRLALIRASVHGKJRI
MPPVALLTFRFKAGILEEKAKIYAEVAVANDIEVGINIEIPAAALADQFAKEV
DFFSITGNDLLOYTMAADRMNEQSYLYQPNPSILRLINNVIKAAHAEKGWAGMCGE
MAGQQTAVPLLVGLDGFDSMSATSLRSLMKKLDFTAKMEEYANRALTECSTMEEV
LELSKKYVNDV
3785..>3991
/gene="gapN"
3785..>3991
/gene="gapN"
/codon_start=1
/transl_table=11
/product="glyceraldehyde 3-phosphate dehydrogenase"
/protein_id="RAPO5991.1"
/db_xref="GI:30027108"
/translation="MAKOYKNVNGEWKTSSENSITIVAPANGEELGSVPAMSAQAEVDE
VYAAKALPAWRALSYAERAAYPs"

ORIGIN
Query Match          74.4%; Score 18.6; DB 1; Length 3991;
Best Local Similarity 84.0%; Pred. No. 2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTGTCGCAA 25
||||| ||||| ||||| ||||| |||||
Db 2097 CAAATCAAGAACTATCGTGCTA 2121

RESULT 53
BS0000014/c
LOCUS
DEFINITION
Pan troglodytes chromosome 22 clone:PTB-065D08, map 22, complete
sequences.
ACCESSION
BS0000014 E0000046
VERSION
BS0000014.1 GI:37537281
KEYWORDS
HTG.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.

REFERENCE
1
The International Chimpanzee Chromosome 22 Consortium.
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429, 382-388 (2004)
2 (bases 1 to 4243)
Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
Fujiyama,A. and Sakaki,Y.
Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.

----- Genome Center
Center: RIKEN Genomic Sciences Center
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project

```

```

Center clone name: PTB-065D08
----- Summary Statistics
Sequencing vector: pUC18,pUC13,prz19R; 100% of reads Chemistry:
Dye-terminator Big Dye and ET; 100% of reads Assembly program:
Phrap; version 0.990329
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.
-----
Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pBACe3.6
The CHORI-251 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/chimpanzee251.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pTARBAC2.1
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKSI45
The PTB22 chimpanzee Fosmid library was prepared from DNA isolated
from cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKSI43
-----
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.
-----
Neighboring clones: RP43-009002(left) and PTB-017F24(right).
Location/Qualifiers
1..4243
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-065D08"
/clone_lib="PTB1 chimpanzee BAC"

FEATURES
source
Query Match          74.4%; Score 18.6; DB 8; Length 44243;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

ORIGIN

```

```

QY 1 CAAATTAAGAGACTATTGTCGCAA 25
Db 24484 CAAATTAAGAGACTATTGTCGCAA 24460

RESULT 54
WPCOMMENT
Sequence split into 5 fragments LOCUS CR735128 Accession CR735128
Fragment Name Begin End
CR735128_0 1 110000
CR735128_1 100001 210000
CR735128_2 200001 310000
CR735128_3 300001 410000
CR735128_4 400001 456610
Continuation (5 of 5) of CR735128 from base 400001 (CR735128 Danio rerio clone DKEY-2010)

Query Match 74.4%; Score 18.6; DB 14; Length 56610;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGTCGCAA 25
Db 29540 CAGATTAAAGAGACTATTCACCCAA 29516

RESULT 55
AC017558/c
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION
AC017558
VERSION
AC017558.1 GI:6554439
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 .58651
AUTHORS
Adams,M. and Venter,J.C.
JOURNAL
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211255 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
Location/Qualifiers
1..58651
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 74.4%; Score 18.6; DB 14; Length 58651;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGTCGCAA 25
Db 55556 CAAATTAAGAGACTATTGTCGCAA 55532

RESULT 56
AC005421/c
LOCUS
DEFINITION
Drosophila melanogaster, chromosome 2L, region 22A1-22A1, P1 clone
DS03601, complete sequence.
ACCESSION
AC005421
VERSION
AC005421.1 GI:3818344

KEYWORDS
HTG.
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 69992)
AUTHORS
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoakins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Masda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
Zieran,L.L. and Rubin,G.M.
TITLE
Sequencing of Drosophila chromosome 2L, region 22A1-22A1
JOURNAL
Unpublished (1998)
REFERENCE
2 (bases 1 to 69992)
AUTHORS
Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
TITLE
Direct Submission
JOURNAL
Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT
On Oct 31, 1998 this sequence version replaced gi:3805728.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
P1 library location: 38-49
FEATURES
source
Location/Qualifiers
1..69992
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strains="Y2: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="22A1-22A1"
/clone="P1 DS03601 (D334)"
/clone_lib="P1 library, partial Sau3A in pAdiosacBII"
/notes="DS03601 (D334) is a partial bridge which extend
from P1 end at bp 1 and extends to bp 69,992, minimally
overlapping DS08613 (D65)."

ORIGIN
Query Match 74.4%; Score 18.6; DB 2; Length 69992;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGTCGCAA 25
Db 68289 CAAATTAAGAGACTATTGTCGCAA 68265

RESULT 57
AC004439/c
LOCUS
DEFINITION
Drosophila melanogaster DNA sequence (P1 DS08613 (D65)), complete
sequence.
ACCESSION
AC004439 AC000837-AC000848 AC001344 AC001345 AC001346 AC001347
AC001836-AC001847 AC002137 AC002888 AC002889 AC002890 AC002891
L81558 L81559 L81560
VERSION
AC004439.1 GI:3342722

```

KEYWORDS	HTG.	JOURNAL	and Salanoubat, M.
SOURCE	Drosophila melanogaster (fruit fly)	REFERENCE	Unpublished
ORGANISM	Drosophila melanogaster	TITLE	2 (bases 1 to 91303)
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	JOURNAL	EU Arabidopsis sequencing project.
AUTHORS	1 (bases 1 to 84055) Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A., Svirskaas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.E.	Submitted (04-FEB-2000)	MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, Bp191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr
TITLE	Sequencing of Drosophila chromosome 2L, region 22A1-22A3	COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> .
JOURNAL	Unpublished (1997)	FEATURES	Location/Qualifiers
AUTHORS	2 (bases 1 to 84055) Celniker, S.E., George, R.A., Galle, R., Svirskaas, R.R., Hoskins, R.A., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.E.	source	1. 91303 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="3" /ecotype="Columbia" 1815..3883 /gene="F24I3.10" join(1815..1871,2260..2448,2531..2839,2912..3145,3239..3883) /gene="F24I3.10" /note="similarity to several hypothetical proteins - different species" /codon_start=1 /product="putative protein" /protein_id="CAB72183.1" /db_xref="GI:6911863" /db_xref="GOA:Q9M1K5" /db_xref="InterPro:IPR001594" /db_xref="UniProt/TREMBL:Q9M1K5" /translation="MANNEKLYOVWRSNKFVCGRLIFGPDASSLVLTLLILGPVWVFVKMTTKMADPTKPNLCIFILCVSWLTILDTLFFLMTSRDGIYPRSPRPETDADPSTTSPSEWVSGRTPNIRIPRVKDVTVNGHTVKVFCDTCLLYRPRASHCISCNVCQVDFDHCPCWGCIGVRNRYFFMFISTTLCTIYVAFSMLNIFQRHMDKISIKWAKISKDLIVLYCFITVWFVGLTIFHSYLCITNTTVENFRYDKKKNYNGKILGNWIEIFLSKIPSPSNKFRSFVKEDYMMVMVETPTSLGSLVSKKIDIEMGGRIVDESGKSYSLPEILNLNVEDLEDDCEDDLKADHSHHHHHQHNEGIIPFPDFFFTNEIGSNKDERNGEESGSSDGTGKRVVRSDEDEKVEGYERNWS 1815..1871 /gene="F24I3.10" /number=1 1872..2259 /gene="F24I3.10" /number=1 2260..2448 /gene="F24I3.10" /number=2 2449..2530 /gene="F24I3.10" /number=2 2531..2839 /gene="F24I3.10" /number=3 2840..2911 /gene="F24I3.10" /number=3 2912..3145 /gene="F24I3.10" /number=4 3146..3238 /gene="F24I3.10" /number=4 3239..3883 /gene="F24I3.10" /number=5 4604..6279
TITLE	Submitted (17-MAR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA	gene	
JOURNAL	On Jul 28, 1998 this sequence version replaced gi:3334915.	CDS	
AUTHORS	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720		
COMMENT	For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site ( <a href="http://fruitfly.berkeley.edu/sequence/">http://fruitfly.berkeley.edu/sequence/</a> ) or send email to <a href="mailto:drosophila@hgsc.lbl.gov">drosophila@hgsc.lbl.gov</a> . Library location: 69-90. Location/Qualifiers		
FEATURES	source	exon	
ORIGIN	1. 84055 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="2L" /map="22A1-22A3" /clone="P1 DS08613 (D65)"	intron	
Query Match	74.4%; Score 18.6; DB 2; Length 84055;	exon	
Best Local Similarity	84.0%; Pred. No. 1.1e+03;	intron	
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	exon	
QY	1 CAATTAAGACACTTTCGTCAA 25 	intron	
Db	1068 CAATTAAGCAACTATTGTCAA 1044	exon	
RESULT 58		intron	
ATF24I3/C		exon	
LOCUS	ATF24I3 91303 bp DNA linear PLN 16-APR-2005	intron	
DEFINITION	Arabidopsis thaliana DNA chromosome 3, BAC clone F24I3.	exon	
ACCESSION	AL138655	intron	
VERSION	AL138655.1 GI:6999905	exon	
KEYWORDS	Arabidopsis thaliana (thale cress)	intron	
ORGANISM	Arabidopsis thaliana	exon	
SOURCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	intron	
REFERENCE	1 Nyakatura, G., Fartmann, B., Dauner, D., Steir, W., Holland, R., Weichselgartner, M., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F.	exon	
AUTHORS		gene	



CDS /gene="F2413.20"  
Join(4604. .5077,5292. .5356,5433. .5760,5838. .5936,  
6016. .6279)  
/gene="F2413.20"  
/note="in the published mRNA sequence a downstream lying  
ATG at position 4712 is used as start codon"  
/codon\_start=1  
/product="leucine zipper-containing protein AT103"  
/protein\_id="CAB72164.1"  
/db\_xref="GI:6911864"  
/db\_xref="GOA:Q9MIK4"  
/db\_xref="UniProt:IPR008434"  
/db\_xref="InterPro:TREMBL:Q9MIK4"  
/translation="MAAEMALVKPISKFSPPKLSNPSPKFLSGRRFSTVIRMSASSPP  
PTTATSKGTKEI0BSLTPRYTDFEEMQLFNTINKNLNEAEFPALLQEF  
KTDVQTHFVRNKEKRAADKLGPLROI FVEFLERSC TAFSGFLLYKELGRRLLKKT  
NPVAAIEFSLSRDEARHAGFLNGLSDFNALDGLFTKARKYTFEPKFFPYATYL  
SEKIGYRITITIRHLKENPFQCYPIFAIFENWCODENRHGDFFSALMKAQPFIND  
WQAKLSRFCLSVVTVMYVNDQRTNFEYEGILNTKFDHVIITETRTTAFIPAV  
LDVNPFPKRLDRMVVSYEKLLAIGETDDASFIKTLKRIPLVTSLSASEILAAVLMPP  
VSSGSVDPAEEFEPNLVY"  
4604. .5077  
/gene="F2413.20"  
/number=1  
5078. .5291  
/gene="F2413.20"  
/number=1  
5292. .5356  
/gene="F2413.20"  
/number=2  
5357. .5432  
/gene="F2413.20"  
/number=2  
5433. .5760  
/gene="F2413.20"  
/number=3  
5761. .5837  
/gene="F2413.20"  
/number=3  
5838. .5936  
/gene="F2413.20"  
/number=4  
5937. .6015  
/gene="F2413.20"  
/number=4  
6016. .6279  
/gene="F2413.20"  
/number=5  
6596. .7654  
/gene="F2413.30"  
complement(Join(6596. .6760,6849. .7109,7367. .7654))  
/gene="F2413.30"  
complement(Join(6596. .6760,6849. .7109,7367. .7654))  
/gene="F2413.30"  
/note="similarity to NODULIN-26 - Glycine max,  
EMBL:X04782"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB72165.1"  
/db\_xref="GI:6911865"  
/db\_xref="GOA:Q9MIK3"  
/db\_xref="UniProt:IPR000425"  
/db\_xref="InterPro:Swiss-Prot:Q9MIK3"  
/translation="MGRIGLVVDLVSEFWI WAGVLVNLVHGVLFGRTPDPSGIV  
RYLFISISMFIYALQATKGGIYNPLTALAAGVSGSPSIFPSVFRIPVVEVIGSIL  
AVKHIIVPEIKGKPLNVAIHGALTGILTFVTLVSMGLTRKIPGSPFMTWVIG  
SLAKLTHILGSLTGGCMNPAVMGWAYARGEHITKEHLLVYVWLGPKVATLLAVNWF  
KVVPKLTREQERPKASE"  
complement(6596. .6760)  
/gene="F2413.30"  
/number=1  
complement(6761. .6848)  
/gene="F2413.30"

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

gene

gene

CDS

exon

intron

/number=1  
complement(6849. .7109)  
/gene="F2413.30"  
/number=2  
complement(7110. .7366)  
/gene="F2413.30"  
/number=2  
complement(7367. .7654)  
/gene="F2413.30"  
/number=3  
8967. .11895  
/gene="F2413.40"  
Join(8967. .10168,10257. .10424,10510. .10675,10756. .10872,  
10961. .11154,11233. .11363,11450. .11522,11607. .11895)  
/gene="F2413.40"  
/note="strong similarity to  
phosphatidylinositol-4-phosphate 5-kinase AtPIP5K1 -  
Arabidopsis thaliana, EMBL:AB005902  
Contains ATP/GTP-binding site motif A (P-loop) AA102-109"  
/codon\_start=1  
/product="phosphatidylinositol-4-phosphate 5-kinase-like  
protein"  
/protein\_id="CAB72166.1"  
/db\_xref="GI:6911866"  
/db\_xref="GOA:Q9MIK2"  
/db\_xref="InterPro:IPR002498"  
/db\_xref="InterPro:IPR003409"  
/db\_xref="UniProt/TREMBL:Q9MIK2"  
/translation="MSKQSCVLKAWETVRKTOQAKKRANSIFGTVSVAPOTDGDDAT  
TTTENDDETSTNRSSIGELSYHAERILPENGDYTGOWYDSFPHGHGKYLWTGCMYIG  
DWYNGKTWGRGKFGWPGATYEGEFKSGTMDGVGLYTGPSGDTYKGQWYNNLKHGHGI  
KRFPANGDYDGEWRGLQEAQGYQWRDGSYYTGWEMKNATICGKGTFFIWDGNRYDGF  
WDDGPPRGNGTFFKADGSGFYVGHWSNDPEEMNGTYYPGDDGSPENDPKDVFNLSEY  
KICSGERVPLPSOKLSVWNSKVEKPRRMSVDGVSVDVRAPEKNNMGTSEGE  
GAADIDSTTRDLDAEMMLEAEFIOSLRSPSPMRLPRAGKQGETTSKGRNYVEL  
MLNLQIGIRHAGVKAOPVVSLLDKHSAEDPKKVTREPGETKYTPHQSSSEFKWKD  
YCPVFRSLRKLKFKVDPADYMLSTICNDALRSLSPGKSGSPFLYTLNDRYMLKTKKK  
SETKVLRLMLAAYNNHRAFENTLIRFYGLHCVLKGPIQKKRVIRVINGNLPCSEYS  
IHRFPDLKSSLGRTTDPSEINSNTILKDLNLFIRLQKAWQOEFIRQVQDKCEF  
LEQERIMDYSLVAGIHFREASVAGELIPSGARTPIGEFEDESAPRLSRADVDQLLSDP  
TWFASIRLGGNNMPARAERTMRRSDCEFLVGEPTEGYEYVVMIFGIIDILQDYDISKK

Query Match 74.4%; Score 18.6; DB 15; Length 91303;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGACACTATTGTCGAA 25  
DB 81795 CAAAAACAGAGACTATTGTCGAA 81771

RESULT 59  
AC002407/c  
LOCUS Human Chromosome X, complete sequence.  
DEFINITION AC002407  
ACCESSION AC002407.1 GI:2323254  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 95506)  
Chen, E., Brownstein, B. H., States, D. J., Schlessinger, D. and  
Mazzarella, R.  
Direct Submission  
Unpublished (1997)  
2 (bases 1 to 95506)  
Brownstein, B. H., States, D. J. and Mazzarella, R.  
Direct Submission  
Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,  
Washington University School of Medicine, 4566 Scott Avenue, St.

COMMENT  
 Louis, MO 63110, USA  
 Current status of this project is available at:  
 'http://genome.wustl.edu/cgm/seq\_projects.html',  
 Submitted by:  
 Ellison Chen,  
 Advanced Center for Genetic Technology,  
 Applied Biosystems Division of Perlin Elmer Corp.,  
 850 Lincoln Center Drive,  
 Foster City, CA 94404 USA  
 e-mail: ellison@genseq.apldbio.com

and

Buddy Brownstein,  
 Center for Genetics in Medicine,  
 Washington University School of Medicine, Box 8232  
 4566 Scott Avenue,  
 St. Louis, MO 63110, USA  
 e-mail: buddy@genetics.wustl.edu

and

David J. States,  
 Institute for Biomedical Computing  
 Washington University in St. Louis  
 700 South Euclid Ave.  
 St. Louis, MO 63108 USA  
 e-mail: states@ibc.wustl.edu.

#### FEATURES

source  
 1..95506  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="bWDXD177"

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 95506;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCAA 25  
 |||||  
 Db 85341 CAATTAAGAGACAATATTTTGCAA 85317

#### RESULT 60

AL451127 108958 bp DNA linear PRI 18-MAY-2005  
 LOCUS Human DNA sequence from clone RP11-12817 on chromosome 9 Contains a  
 DEFINITION novel gene, complete sequence.

ACCESSION AL451127 AC025854

VERSION AL451127.6 GI:14018295

KEYWORDS HTG.

SOURCE Homo sapiens (human)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 108958)

#### REFERENCE

AUTHORS Brown, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

#### COMMENT

Clone requests: clonerequest@sanger.ac.uk  
 On or before May 15, 2001 this sequence version replaced  
 gi:8077006, gi:13751503.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human

chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr9  
 RP11-12817 is from the library RPCI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk  
 -----

Draft Sequence Produced by Whitehead Institute/MIT Center for  
 Genome Research, 320 Charles Street,  
 Cambridge, MA 02141, USA  
 http://www-seq.wi.mit.edu

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one subclone; and the assembly was confirmed by restriction digest,  
 except on the rare occasion of the clone being a YAC.

#### FEATURES

Location/Qualifiers  
 1..108958  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-12817"  
 /clone\_lib="RPCI-11.1"  
 100  
 misc\_feature  
 /note="Clone right\_end: RP11-404E6"  
 join(92855..93084,97109..97192,98429..98738)  
 /locus\_tag="RP11-12817.1-001"  
 join(92855..93084,97109..97192,98429..98738)  
 /locus\_tag="RP11-12817.1-001"  
 /note="match: ESTs: BM718494.1"  
 108958  
 misc\_feature  
 /note="Clone\_right\_end: RP11-12817"

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 108958;  
 Best Local Similarity 84.0%; Pred. No. 1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCAA 25  
 |||||  
 Db 76723 CAATTAAGAGACTTTTGTGTTAA 76747

Search completed: January 27, 2006, 23:14:00  
 Job time : 997.18 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 784.944 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaacagg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

GenBank:

1: gb\_ba:

2: gb\_in:

3: gb\_env:

4: gb\_cm:

5: gb\_ov:

6: gb\_pat:

7: gb\_ph:

8: gb\_pr:

9: gb\_ro:

10: gb\_sts:

11: gb\_sy:

12: gb\_uni:

13: gb\_vi:

14: gb\_htg:

15: gb\_pl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	CS109173 Sequence
2	20	100.0	1731	6	CQ647508 Sequence
3	20	100.0	1734	6	AX608629 Sequence
4	20	100.0	21781	1	AE014229 Streptoco
5	20	100.0	29072	6	AX602188 Sequence
6	20	100.0	167050	1	SAG766847 Streptoco
7	20	100.0	349980	6	CQ655069 Sequence
8	20	100.0	349980	6	AX954529 Sequence
9	18.4	92.0	396	13	HIVAJ8558 Human imm
10	18.4	92.0	396	13	HIVAJ8564 Human imm
11	18.4	92.0	396	13	HIVAJ8568 Human imm
12	18.4	92.0	396	13	HIVAJ8582 Human imm
13	18.4	92.0	396	13	HIVAJ8589 Human imm
14	18.4	92.0	396	13	HIVAJ8639 Human imm
15	18.4	92.0	396	13	HIVAJ8642 Human imm
16	18.4	92.0	396	13	HIVAJ8646 Human imm
17	18.4	92.0	396	13	HIVAJ8656 Human imm
18	18.4	92.0	398	13	HIVAJ8543 Human imm

19	18.4	92.0	445	13	AF395561	AF395561 HIV-2 iso
20	18.4	92.0	708	13	HIVFAP	L33089 Human immun
21	18.4	92.0	9472	13	HIV2D194	J04542 Human immun
22	18.4	92.0	9472	13	HIVD194	X52223 Human immun
23	18.4	92.0	9473	6	A09995	A09995 HIV-2 (HIV-
24	18.4	92.0	10312	13	HIV2U38293	U38293 Human immun
25	18.4	92.0	10359	13	HIV2BEN	M30502 Human immun
26	18	90.0	37676	8	AY341245	AY341245 Homo sapi
27	18	90.0	110000	15	CR382125_04	Continuation (5 of
28	18	90.0	114519	8	AL157813	AL157813 Human DNA
29	18	90.0	194045	9	AL833774	AL833774 Mouse DNA
30	18	90.0	204712	14	AC134627	AC134627 Rattus no
31	18	90.0	276881	14	AC095481	AC095481 Rattus no
32	17.4	87.0	491	10	G52355	G52355 SHOC-B5070
33	17.4	87.0	549	10	G90172	G90172 S10P6031RB
34	17.4	87.0	723	10	BV558070	BV558070 qmm55f08
35	17.4	87.0	737	10	BV586373	BV586373 G591P6318
36	17.4	87.0	800	10	BV576492	BV576492 G591P6083
37	17.4	87.0	800	10	BV628207	BV628207 S217P6230
38	17.4	87.0	806	10	BV594335	BV594335 taw08f01
39	17.4	87.0	2081	5	BC063923	BC063923 Xenopus t
40	17.4	87.0	2286	8	AK123724	AK123724 Homo sapi
41	17.4	87.0	2303	2	AF401640	AF401640 Naegleria
42	17.4	87.0	2312	2	AY075353	AY075353 Drosophill
43	17.4	87.0	2345	2	AY089251	AY089251 Drosophill
44	17.4	87.0	4310	2	AF051933	AF051933 Drosophill
45	17.4	87.0	4911	6	CQ585744	CQ585744 Sequence
46	17.4	87.0	7886	6	CQ585743	CQ585743 Sequence
47	17.4	87.0	27103	14	AL353656	AL353656 Homo sapi
48	17.4	87.0	37170	8	HSB12389	Z68224 Human DNA
49	17.4	87.0	39601	15	AC079678	AC079678 Arabidops
50	17.4	87.0	50821	15	AC020579	AC020579 Arabidops
51	17.4	87.0	51405	14	AC019512	AC019512 Drosophill
52	17.4	87.0	54288	14	AC100240	AC100240 Mus muscu
53	17.4	87.0	63540	14	AC101468	AC101468 Mus muscu
54	17.4	87.0	65756	8	AL590286	AL590286 Human DNA
55	17.4	87.0	69817	15	AC003096	AC003096 Arabidops
56	17.4	87.0	84987	8	AL662897	AL662897 Human DNA
57	17.4	87.0	101526	14	AC106751	AC106751 Homo sapi
58	17.4	87.0	104632	8	AL136124	AL136124 Human DNA
59	17.4	87.0	110000	1	CR543861	Continuation (33 o
60	17.4	87.0	110000	14	CR848719_2	Continuation (3 of

ALIGNMENTS

RESULT 1

CS109173

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

CS109173 Sequence 4 from Patent EP1541697.  
CS109173  
CS109173.1 GI:68148022

linear PAT 22-JUN-2005

synthetic construct  
other sequences; artificial sequences.

1  
Unl.J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.  
Detection of group B streptococcus  
Patent: EP 1541697-A 4 15-JUN-2005;  
Mayo Foundation for Medical Education and Research (US)

Location/Qualifiers  
1. ..20  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Oligonucleotide"

Query Match 100.0%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 1 CAAGTAAATGCAGAAACAGG 20

RESULT 2
CQ647508 1731 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 4465 from Patent WO0234771.
DEFINITION
ACCESSION CQ647508
VERSION CQ647508.1 GI:41683480
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 292 CAAGTAAATGCAGAAACAGG 311

RESULT 3
AX608629 1734 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 6558 from Patent WO02092818.
DEFINITION
ACCESSION AX608629
VERSION AX608629.1 GI:28404206
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Glaser,P., Rueniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,B., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 6558 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1..1734
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 292 CAAGTAAATGCAGAAACAGG 311

RESULT 4
AE014229 21781 bp DNA linear BCT 23-SEP-2002
LOCUS Streptococcus agalactiae 2603V/R section 39 of 100 of the complete
DEFINITION genome.
ACCESSION AE014229 AE009948
VERSION AE014229.1 GI:22533831
KEYWORDS
SOURCE Streptococcus agalactiae 2603V/R
ORGANISM Streptococcus agalactiae 2603V/R
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Bretttoni,C., Galli,G.,
Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
TITLE Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
12200547
2 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Bretttoni,C., Galli,G.,
Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
Direct Submission
Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..21781
/organism="Streptococcus agalactiae 2603V/R"
/mol_type="genomic DNA"
/strain="2603V/R"
/serotype="V"
/db_xref="taxon:208435"
103..348
/gene="SAG0813"
103..348
/gene="SAG0813"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM99700.1"
/db_xref="GI:22533832"
/transl_table="MNQINHCFKSIMVLGVGTQMKKEFIITGRMVARVMTLVCLCMV
SNLVLDMLDIIEINEIVFCVLVASIAGLEFLQKERE"
387..674
/gene="SAG0814"
387..674
/gene="SAG0814"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM99701.1"
/db_xref="GI:22533833"
/transl_table="MTYIGVPKEQKALSRRGKQLEKINQGGKTLMSQIMQIDLVGSTCM
IYLEDGSOTHPYEILKVDVDEWQVIRRNLRKSRKEVRIIVRIHDIKELRPVK"
719..934
/gene

```

```
/gene="SAG0815"
/719..934
/gene="SAG0815"
/notes="identified by match to PFAM protein family HMM
PF00381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/Ci family"
/protein_id="AAM99702.1"
/db_xref="GI:22533834"
/translation="WGLILKNRLKELRARDGINQTELAKLTGVSRTISLIERGEVTP
SVIAWKISQVFKESLENVFLSVSME"
934..1695
/gene="SAG0816"
/gene="SAG0816"
/notes="similar to OMN1:NTL01SPL1088, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99703.1"
/db_xref="GI:22533835"
/translation="MMKNFIKQSDTTFVRIIISLLIGGFIGAILSGVGFIIFGQDK
YLSINIVQFLWRSIVIVITALLSFLIYIQIKYQKRVFNVDSESEIYRQINLR
HSYGMTFVGISIVLNTLVFNKLNIFDSDVTLPIYDLSLLFLLGLHIIYFLKVI
RNIRGFKMTVAPTLKELKNVLQDLAELESNYKMFIDVWNLSGFIPTIYFVLFFI
SPVQKVEIVAIITTSIHIIILIKSLKAARHFR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/notes="similar to OMN1:NTL01SPL0446, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99704.1"
/db_xref="GI:22533836"
/translation="MSKNNYTCIIETAIFAALAMALSIMPASWFTPSFGAIPILIL
FALRRKTKYGLFAGLILGLHFLVLSKYVLSQVFEIYLAFISMGLAGVPSAKFD
ALLSSSKTKALSALGAILATLVRYVWHYIAGVIFWASYAPKMSATLISUSVNGTA
GLILFLFVVISIILVISYPSFLPKK"
complement(2584..3543)
/gene="nrdf-2"
/notes="synonym: SAG0818"
complement(2584..3543)
/gene="nrdf-2"
/notes="identified by match to PFAM protein family HMM
PF00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AAM99705.1"
/db_xref="GI:22533837"
/translation="WTTYEAINWNEIEDVSKTEKLTEQFWLDTRIPLSLDLDDM
RKLSAQEKLVGFGGLTLLDTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVRAK
SYSISFTLNKSEIEIEFWTNNEFLQEKARINDIYANGNALQKVKASTYLETFL
FYSGFTPLLYLGNKLANVAEIKLIIRDESVMHTYIGYKQLGFNLPEDQENFR
DMWYDLLYLYNEEKYTKLYDGVGMTEVMTFLRYNANKALMNLGQDPLFPDPTAND
VNFIVMNGISTGTSNHDFFSQVNGYLLGSVEAMHDDDDNYNGL"
complement(3746..5905)
/gene="nrde-2"
/notes="synonym: SAG0819"
complement(3746..5905)
/gene="nrde-2"
/notes="identified by match to PFAM protein family HMM
PF00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
```

```
/protein_id="AAM99706.1"
/db_xref="GI:22533838"
/translation="MSLKNIGDVSYFRLNNEINRPVNGQIPLHKDKALTAFFKENVQ
PNSKAFDSITDKIAYLLKYDLEEAFLNKYRPEFIEELSTKLFDKFRKPSMAAYKF
YQOVALKTNDGEYLYLESTEDRVLFNALYFADGDEBLATDLALEMISQRPATP9FLN
AGSRGSELVSCFLIQVTDMMNAIGRSINSAQLSRIGGVGIGISLNLREAGAPIKGF
AGASGVYVPMKLFEDSPSYSNQLGQROGAGVYLDVDFHPDIIISFLTKKENADEKVR
VATLSGIVTPDKFYELARNQEMTLFSPYSIEREYGVFFSIDITEKDELVANPNI
TKTKINARDLETSIKLQOESGYPIINIDTANRTPVDGKIIMSLNLCSEILQVQKPS
LINDAQEYLEMTDISCNLSTVNLMNTSPDGKSIKTMTRALTFTVDSSEINAEVPT
IKNGNAQHTFGLGAGLHSLYAKNHIYGPSIESIEFTDIYFMLNMYTWLVSNNIAR
ERQTTVGFPEKSYADGYFDKYVSGKFVQSDKVSLFANHFIIPEAKDWENLRVAM
KDGLYHQNLAVAPNGSISYINDCSASIHPIQRIEERQEKIKGIYYIPANGLATDTI
PYTTSAYDMRMKRVLDVAAATEHVDQGLSMTLFLRSELKELYEWKTESKQTTTRDLS
ILRNAYFNKGVKSYIIRTFDTHDGVSEVGANQCESVI"
complement(5983..6207)
/gene="nrde"
/notes="synonym: SAG0820"
complement(5983..6207)
/gene="nrde"
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2,
nrde-redoxin"
/protein_id="AAM99707.1"
/db_xref="GI:22533839"
/translation="MAAITVFSKNNCMQCKMTKKFLDQHGADPFEINIDEKPEKIEYV
KNLGSAAAPVIEAGNVVFSGFQPSKLKELV"
6590..6853
/gene="ptsh"
/notes="synonym: SAG0821"
6590..6853
/gene="ptsh"
/notes="identified by match to PFAM protein family HMM
PF00381"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAM99708.1"
/db_xref="GI:22533840"
/translation="WASKDPFHVAETGIHARPATLIVQTASKFASDITLDYKGAENVL
KSIQVMSLGVCGQADVTISAGGADDDAIAAIEETMTKEGLA"
6858..8591
/gene="ptsh"
/notes="synonym: SAG0822"
6858..8591
/gene="ptsh"
/notes="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAM99709.1"
/db_xref="GI:22533841"
/translation="MTEMLUKGIAASGVAVAKAYLLVQDLSPEFTVVDNABEARL
DVALAQSDQLSVIREKAGEEAAVDFADHMLVLSDPMINQIKETIRAKOVNAB
TGLKEVTDMFITIFEGMEDNPYQERAADIRDVAKRVLAHLLGVKLFPNATINEESIV
IAHDLTPSSTAQLNKRQFVKAFVFNIGRTSHSALMARTLEIAAVLGTNDITERVQDQ
```

Query Match 100.0%; Score 20; DB 1; Length 21781;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAGG 20  
|||||  
DB 7149 CAAGTAATGCAGAAACAGG 7168

RESULT 5  
AX602188 29072 bp DNA linear PAT 17-FEB-2003  
LOCUS  
DEFINITION Sequence 117 from Patent WO02092818.  
ACCESSION AX602188



/product="Unknown"  
/protein\_id="CAD46414.1"  
/db\_xref="GI:23095204"  
/db\_xref="GOA:Q8E657"  
/db\_xref="InterPro:IPR002180"  
/db\_xref="UniProt/Swiss-Prot:Q8E657"  
/translation="MTIEGQLVANEMKIGIVVSRFNEIITSKLLSGAVDGLLRHGV  
EEDIDIVVPGAFIPIYMARKWALYKDIAICLVVIGKSTHDYDVGVNEVTKGIGH  
LNSQSDPHIFGVLTDDTNEQAIERAGTKAGNKGVCALSAIEMVNLDKLRER"  
complement(3814..5315)  
/gene="lysS"  
complement(3814..5304)  
/gene="lysS"  
/codon\_start=1  
/transl\_table=11  
/product="lysoyl-tRNA synthetase"  
/protein\_id="CAD46415.1"  
/db\_xref="GI:23095205"  
/db\_xref="GOA:Q8E656"  
/db\_xref="InterPro:IPR002313"  
/db\_xref="InterPro:IPR004364"  
/db\_xref="InterPro:IPR004365"  
/db\_xref="InterPro:IPR006195"  
/db\_xref="InterPro:IPR008994"  
/db\_xref="UniProt/Swiss-Prot:Q8E656"  
/translation="MNSQIEELNDQIVRREKMAALTEQGDIPFGKRFERTATSGQL  
NEKYADSKEDLHDIRETATIGRLMTKRGKGVGAHIDQREGQIQIVRKDSVCEE  
NYEIPKADLGDPLGVGGVMRTDMGELSIAKATHILSKALPLPEKFGHLDITFI  
YRKRLDLISNRDSRFVTRKSIIEIRRFMSNGFLEVTTPVLDNHNAGGASARFPI  
THNQADIDMLVATIELHLKRLIVGGMERVYIGIRFNEGMDATHNPEFTSEIAYQ  
AVAYQDIDLITGIIQHVTKVKGDPYINQGTIEIKNEPFKRVMHVDVKEITGID  
FWKEMLEBAQAQKVPLEKHTTIVGHHINAFPEEFVEDTLIQPTFVFGHPVSV  
PLAKNDTPRFTDRELEFIMTKYANATFELNDPIDQLSRPEAQASAKELGDEATG  
VDYDYEALEYGNPPYGGIGIDRLCMLLTDTTIRDVLLFFPTMK"  
complement(5310..5315)  
/gene="lysS"  
5467..6411  
/gene="gbs0772"  
5467..5472  
/gene="gbs0772"  
5479..6381  
/gene="gbs0772"  
/note="Similar to unknown proteins"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46416.1"  
/db\_xref="GI:23095206"  
/db\_xref="GOA:Q8E655"  
/db\_xref="InterPro:IPR005833"  
/db\_xref="InterPro:IPR005834"  
/db\_xref="InterPro:IPR006439"  
/db\_xref="UniProt/TREMBL:Q8E655"  
/translation="MITSIVFDVDDTIYDQAPYIAVEKCFDFDMSAINQAIYRFR  
HSDIGDPRLVMAEWTTEYFRFWRKCTELLFEGVEIDEATGIYFOBIYRHELENTIM  
LDENWTLTKSKNVPWGIITNGPTEHQLKVKKGLDYDVPKRVIVSQAGFKQPK  
EKEIFNLAAQDFMNPSTLYVGDSDYNDINGAFNGSWHSMWFNHRGRSLKPKIKPVY  
DVAIDNPEQLFAGVLDLFDLPDNKFIPIIDNKNPNVLEMLNGLMMAERKLESNMS  
VDKVVILLRTAQEKVLRMKYAR"  
6390..6411  
/gene="gbs0772"  
complement(6417..7069)  
/gene="gbs0773"  
complement(6417..7058)  
/gene="gbs0773"  
/note="Similar to putative phosphoglycerate mutase"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46417.1"  
/db\_xref="GI:23095207"  
/db\_xref="GOA:Q8E654"  
/db\_xref="InterPro:IPR001345"

terminator

gene

CDS

/db\_xref="UniProt/TREMBL:Q8E654"  
/translation="MKFYLVHRHGKTMNLEGRFQAGNGSPILLEAIELEBELCOYLS  
SIHFDAVYSSDLGRADTVNILDANSCKPEIHYTPQLREWALGTLEGCKIATMQAIY  
PRMTAFYQNPLQFKHDMFGAESLYQTHRVESFLRSLASQYDKVLIYVGHANLTAS  
IRSLIGYOYGLHYKDKLDNASLTIETHDFKDFNCLTWNDKSYLRQREVKMTH"  
complement(7064..7069)  
/gene="gbs0773"  
complement(7080..7553)  
/gene="gbs0774"  
complement(7080..7553)  
/gene="gbs0774"  
/note="Similar to transcriptional regulator"  
/codon\_start=1  
/transl\_table=11

RBS

gene

CDS

Query Match 100.0%; Score 20; DB 1; Length 167050;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20

Db 79953 CAAGTAAATGCAGAACAGG 79972

RESULT 7

CQ655069 349980 bp DNA linear PAT 02-FEB-2004  
LOCUS  
Sequence 12026 from Patent WO0234771.  
DEFINITION  
ACCESSION CQ655069  
VERSION CQ655069.1 GI:41687946  
KEYWORDS  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE

AUTHORS

Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.  
and Tettelin,H.

TITLE

Nucleic acids and proteins from streptococcus groups a & b

JOURNAL

Patent: WO 0234771-A 12026 02-MAY-2002;

Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES

source

1..349980  
/organism="Streptococcus agalactiae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1311"  
/note="seq 10967 to big: splitted in  
seq 10.967, from 0.000.001 to 0.349.980  
seq 12.025, from 0.300.001 to 0.649.980  
seq 12.026, from 0.600.001 to 0.949.980  
seq 12.027, from 0.900.001 to 1.249.980  
seq 12.028, from 1.200.001 to 1.549.980  
seq 12.029, from 1.500.001 to 1.849.980  
seq 12.030, from 1.800.001 to 2.149.980  
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20

Db 191730 CAAGTAAATGCAGAACAGG 191749

RESULT 8

AX954529

LOCUS AX954529 349980 bp DNA linear PAT 08-JAN-2004  
DEFINITION Sequence 1375 from Patent WO03033306.  
ACCESSION AX954529  
VERSION AX954529.1 GI:40783902  
KEYWORDS

```

SOURCE      Streptococcus pyogenes
ORGANISM    Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS     Telford, J., Maignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.
            and Tettelin, H.
TITLE       Nucleic acids and proteins from streptococcus groups a b
JOURNAL     Patent: WO 03093306-A 1375 13-NOV-2003;
            Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source      Location/Qualifiers
            1..349980
              /organism="Streptococcus pyogenes"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1314"
              /note="seq 1373 to long, original length: 2.160.266
                replaced by
                seq 1373: from 0.000.001 to 0.349.980
                seq 1374: from 0.300.001 to 0.649.980
                seq 1375: from 0.600.001 to 0.949.980
                seq 1376: from 0.900.001 to 1.249.980
                seq 1377: from 1.200.001 to 1.549.980
                seq 1378: from 1.500.001 to 1.849.980
                seq 1379: from 1.800.001 to 2.149.980
                seq 1380: from 2.100.001 to 2.160.266"

ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
        |||||
Db      220517 CAAGTAAATGCAGAAACAGG 220536

RESULT 9
HIVAJ8558
LOCUS      HIVAJ8558             396 bp      DNA      linear      VRL 15-APR-2005
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1033.
ACCESSION  AJ008558
VERSION     AJ008558.1 GI:3286872
KEYWORDS   DNA polymerase; pol gene.
SOURCE     Human immunodeficiency virus 2 (HIV-2)
ORGANISM   Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE
AUTHORS     Xiang, Z., Ariyoshi, K., Wilkins, A., Dias, F., Whittle, H. and
            Breuer, J.
TITLE       HIV type 2 pathogenicity is not related to subtype in rural Guinea
JOURNAL     AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
PUBMED      9100992
REFERENCE
AUTHORS     Grassly, N., Xiang, Z., Ariyoshi, K., Aaby, P., Jensen, H., Dias, F.,
            Whittle, H. and Breuer, J.
TITLE       Correlation of HIV-2 genotype with progression to AIDS in vivo
JOURNAL     Unpublished
REFERENCE
AUTHORS     Xiang, Z.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
            Department of Virology, Queen Mary Westfield, Turner Street,
            London, E1 2AD, UK
FEATURES
source      Location/Qualifiers
            1..396
              /organism="Human immunodeficiency virus 2"
              /mol_type="genomic DNA"
              /strain="A"
              /isolate="b1033"
              /db_xref="taxon:11709"
              <1..>396
            gene

SOURCE      Streptococcus pyogenes
ORGANISM    Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS     Telford, J., Maignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.
            and Tettelin, H.
TITLE       Nucleic acids and proteins from streptococcus groups a b
JOURNAL     Patent: WO 03093306-A 1375 13-NOV-2003;
            Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source      Location/Qualifiers
            1..349980
              /organism="Streptococcus pyogenes"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1314"
              /note="seq 1373 to long, original length: 2.160.266
                replaced by
                seq 1373: from 0.000.001 to 0.349.980
                seq 1374: from 0.300.001 to 0.649.980
                seq 1375: from 0.600.001 to 0.949.980
                seq 1376: from 0.900.001 to 1.249.980
                seq 1377: from 1.200.001 to 1.549.980
                seq 1378: from 1.500.001 to 1.849.980
                seq 1379: from 1.800.001 to 2.149.980
                seq 1380: from 2.100.001 to 2.160.266"

ORIGIN
Query Match      92.0%; Score 18.4; DB 13; Length 396;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
        |||||
Db      194 CAAGTAAATGCAGAAATAGG 213

RESULT 10
HIVAJ8564
LOCUS      HIVAJ8564             396 bp      DNA      linear      VRL 15-APR-2005
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1048.
ACCESSION  AJ008564
VERSION     AJ008564.1 GI:3286884
KEYWORDS   DNA polymerase; pol gene.
SOURCE     Human immunodeficiency virus 2 (HIV-2)
ORGANISM   Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE
AUTHORS     Xiang, Z., Ariyoshi, K., Wilkins, A., Dias, F., Whittle, H. and
            Breuer, J.
TITLE       HIV type 2 pathogenicity is not related to subtype in rural Guinea
JOURNAL     AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
PUBMED      9100992
REFERENCE
AUTHORS     Grassly, N., Xiang, Z., Ariyoshi, K., Aaby, P., Jensen, H., Dias, F.,
            Whittle, H. and Breuer, J.
TITLE       Correlation of HIV-2 genotype with progression to AIDS in vivo
JOURNAL     Unpublished
REFERENCE
AUTHORS     Xiang, Z.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
            Department of Virology, Queen Mary Westfield, Turner Street,
            London, E1 2AD, UK
FEATURES
source      Location/Qualifiers
            1..396
              /organism="Human immunodeficiency virus 2"
              /mol_type="genomic DNA"
              /strain="A"
              /isolate="b1048"
              /db_xref="taxon:11709"
              <1..>396
            gene

SOURCE      Streptococcus pyogenes
ORGANISM    Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS     Telford, J., Maignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.
            and Tettelin, H.
TITLE       Nucleic acids and proteins from streptococcus groups a b
JOURNAL     Patent: WO 03093306-A 1375 13-NOV-2003;
            Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source      Location/Qualifiers
            1..349980
              /organism="Streptococcus pyogenes"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1314"
              /note="seq 1373 to long, original length: 2.160.266
                replaced by
                seq 1373: from 0.000.001 to 0.349.980
                seq 1374: from 0.300.001 to 0.649.980
                seq 1375: from 0.600.001 to 0.949.980
                seq 1376: from 0.900.001 to 1.249.980
                seq 1377: from 1.200.001 to 1.549.980
                seq 1378: from 1.500.001 to 1.849.980
                seq 1379: from 1.800.001 to 2.149.980
                seq 1380: from 2.100.001 to 2.160.266"

ORIGIN
Query Match      92.0%; Score 18.4; DB 13; Length 396;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
        |||||
Db      194 CAAGTAAATGCAGAAATAGG 213

RESULT 10
HIVAJ8564
LOCUS      HIVAJ8564             396 bp      DNA      linear      VRL 15-APR-2005
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1048.
ACCESSION  AJ008564
VERSION     AJ008564.1 GI:3286884
KEYWORDS   DNA polymerase; pol gene.
SOURCE     Human immunodeficiency virus 2 (HIV-2)
ORGANISM   Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE
AUTHORS     Xiang, Z., Ariyoshi, K., Wilkins, A., Dias, F., Whittle, H. and
            Breuer, J.
TITLE       HIV type 2 pathogenicity is not related to subtype in rural Guinea
JOURNAL     AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
PUBMED      9100992
REFERENCE
AUTHORS     Grassly, N., Xiang, Z., Ariyoshi, K., Aaby, P., Jensen, H., Dias, F.,
            Whittle, H. and Breuer, J.
TITLE       Correlation of HIV-2 genotype with progression to AIDS in vivo
JOURNAL     Unpublished
REFERENCE
AUTHORS     Xiang, Z.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
            Department of Virology, Queen Mary Westfield, Turner Street,
            London, E1 2AD, UK
FEATURES
source      Location/Qualifiers
            1..396
              /organism="Human immunodeficiency virus 2"
              /mol_type="genomic DNA"
              /strain="A"
              /isolate="b1048"
              /db_xref="taxon:11709"
              <1..>396
            gene

```



/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQIVNTCAQCKQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFVEAEVIP  
QESGRQTALFLKLASRWPIITHLTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||

Db 194 CAAGTAAATGCAGAAATAGG 213  
|||||

## RESULT 11

HIVAJ8568  
LOCUS HIVAJ8568 396 bp DNA linear VRL 15-APR-2005  
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1053.

ACCESSION AJ008568  
VERSION AJ008568.1 GI:3286892  
KEYWORDS DNA polymerase; pol gene.  
SOURCE Human immunodeficiency virus 2 (HIV-2)  
ORGANISM Human immunodeficiency virus 2

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

REFERENCE 1  
AUTHORS Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and  
Breuer,J.

TITLE HIV type 2 pathogenicity is not related to subtype in rural Guinea

JOURNAL AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)

## REFERENCE 2

Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,  
Whittle,H. and Breuer,J.

TITLE Correlation of HIV-2 genotype with progression to AID in vivo

REFERENCE 3 (bases 1 to 396)

Xiang,Z.

Direct Submission

TITLE Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,  
London, E1 2AD, UK

## FEATURES

source

1..396  
Location/Qualifiers  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="A"  
/isolate="b1053"  
/db\_xref="taxon:11709"

gene

<1..>396  
/gene="pol"

CDS

<1..>396  
/gene="pol"  
/codon\_start=2  
/product="DNA polymerase"  
/protein\_id="CA08032.1"  
/db\_xref="GI:3286893"  
/db\_xref="GOA:Q9YQL5"  
/db\_xref="InterPro:IPR001584"  
/db\_xref="InterPro:IPR003308"  
/db\_xref="UniProt/TREMBL:Q9YQL5"  
/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQIVNTCAQCKQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFVEAEVIP  
QESGRQTALFLKLASRWPIITHLTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||

Db 194 CAAGTAAATGCAGAAATAGG 213

## RESULT 12

HIVAJ8582

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human immunodeficiency virus type 2 (HIV-2)

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

REFERENCE 1

AUTHORS Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and  
Breuer,J.

TITLE HIV type 2 pathogenicity is not related to subtype in rural Guinea

JOURNAL AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)

REFERENCE 2

AUTHORS Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,  
Whittle,H. and Breuer,J.

TITLE Correlation of HIV-2 genotype with progression to AID in vivo

REFERENCE 3 (bases 1 to 396)

Xiang,Z.

Direct Submission

TITLE Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,  
London, E1 2AD, UK

FEATURES

source

1..396  
Location/Qualifiers  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="A"  
/isolate="b1093"  
/db\_xref="taxon:11709"

gene

<1..>396  
/gene="pol"

CDS

<1..>396  
/gene="pol"  
/codon\_start=2  
/product="DNA polymerase"  
/protein\_id="CA08046.1"  
/db\_xref="GI:3286921"  
/db\_xref="GOA:Q9YQK6"  
/db\_xref="InterPro:IPR001584"  
/db\_xref="InterPro:IPR003308"  
/db\_xref="UniProt/TREMBL:Q9YQK6"  
/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQIVNTCAQCKQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFVEAEVIP  
QESGRQTALFLKLASRWPIITHLTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||

Db 194 CAAGTAAATGCAGAAATAGG 213  
|||||

## RESULT 13

HIVAJ8589

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

Human immunodeficiency virus 2 (HIV-2)

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

HIVAJ8589 396 bp DNA linear VRL 15-APR-2005  
Human immunodeficiency virus type 2 pol gene, isolate b1123.

ACCESSION AJ008589  
VERSION AJ008589.1 GI:3286934  
KEYWORDS DNA polymerase; pol gene.  
SOURCE Human immunodeficiency virus 2 (HIV-2)



```

source
1. .396
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/strain="A"
/isolates="b1244"
/db_xref="taxon:11709"
<1..>396
/gene="pol"
<1..>396
/gene="pol"
/codon_start=2
/product="DNA polymerase"
/protein_id="CAA08106.1"
/db_xref="GI:3287041"
/db_xref="GOA:O9YOH6"
/db_xref="InterPro:IPR001584"
/db_xref="InterPro:IPR003308"
/db_xref="UniProt/TREMBL:Q9YOH6"
/translation="DHLVSGQIRQVLFLEKIEPAQBEHEKYHSNVKELSHKFGLPKLV
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIETAEVIP
QESGRQTALFLKLASRWPIIHLHTDNGA"

```

## ORIGIN

```

Query Match      92.0%; Score 18.4; DB 13; Length 396;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    |||||
Db 194 CAAGTAAATGCAGAAATAGG 213

```

```

RESULT 16
HIVAJ8646
LOCUS
DEFINITION
HIV type 2 pathogenicity is not related to subtype in rural Guinea
ACCESSION
AJ008646
VERSION
AJ008646.1 GI:3287048
KEYWORDS
DNA polymerase; pol gene.
SOURCE
Human immunodeficiency virus 2 (HIV-2)
ORGANISM
Human immunodeficiency virus 2
Viruses; Retro-transcribing viruses; Retroviridae;
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

REFERENCE
1 Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and
Breuer,J.
HIV type 2 pathogenicity is not related to subtype in rural Guinea
Bissau
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
9100992
REFERENCE
2
Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,
Whittle,H. and Breuer,J.
Correlation of HIV-2 genotype with progression to AID in vivo
Unpublished
Xiang,Z.
REFERENCE
3 (bases 1 to 396)
Xiang,Z.
Direct Submission
Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
Department of Virology, Queen Mary Westfield, Turner Street,
London, E1 2AD, UK
9100992
REFERENCE
2
Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,
Whittle,H. and Breuer,J.
Correlation of HIV-2 genotype with progression to AID in vivo
Unpublished
Xiang,Z.
REFERENCE
3 (bases 1 to 396)
Xiang,Z.
Direct Submission
Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
Department of Virology, Queen Mary Westfield, Turner Street,
London, E1 2AD, UK

```

```

FEATURES
source
1. .396
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/strain="A"
/isolates="b1253"
/db_xref="taxon:11709"
<1..>396
/gene="pol"
<1..>396
/gene="pol"
/codon_start=2

```

```

gene
CDS

```

```

/product="DNA polymerase"
/protein_id="CAA08110.1"
/db_xref="GI:3287049"
/db_xref="GOA:O77C38"
/db_xref="InterPro:IPR001584"
/db_xref="InterPro:IPR003308"
/db_xref="UniProt/TREMBL:Q77C38"
/translation="DHLVSGQIRQVLFLEKIEPAQBEHEKYHSNVRELSHKFGLPKLV
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIETAEVIP
QESGRQTALFLKLASRWPIIHLHTDNGA"

```

## ORIGIN

```

Query Match      92.0%; Score 18.4; DB 13; Length 396;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    |||||
Db 194 CAAGTAAATGCAGAAATAGG 213

```

## RESULT 17

```

HIVAJ8656
LOCUS
DEFINITION
HIV type 2 pathogenicity is not related to subtype in rural Guinea
ACCESSION
AJ008656
VERSION
AJ008656.1 GI:3287068
KEYWORDS
DNA polymerase; pol gene.
SOURCE
Human immunodeficiency virus 2 (HIV-2)
ORGANISM
Human immunodeficiency virus 2
Viruses; Retro-transcribing viruses; Retroviridae;
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

REFERENCE
1 Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and
Breuer,J.
HIV type 2 pathogenicity is not related to subtype in rural Guinea
Bissau
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
9100992
REFERENCE
2
Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,
Whittle,H. and Breuer,J.
Correlation of HIV-2 genotype with progression to AID in vivo
Unpublished
Xiang,Z.
REFERENCE
3 (bases 1 to 396)
Xiang,Z.
Direct Submission
Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,
Department of Virology, Queen Mary Westfield, Turner Street,
London, E1 2AD, UK

```

## FEATURES

```

source
1. .396
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/strain="A"
/isolates="b1285"
/db_xref="taxon:11709"
<1..>396
/gene="pol"
<1..>396
/gene="pol"
/codon_start=2
/product="DNA polymerase"
/protein_id="CAA08120.1"
/db_xref="GI:3287069"
/db_xref="GOA:O9YJ02"
/db_xref="InterPro:IPR001584"
/db_xref="InterPro:IPR003308"
/db_xref="UniProt/TREMBL:Q9YJ02"
/translation="DHLVSGQIRQVLFLEKIEPAQBEHEKYHSNVRELSHKFGLPKLV
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIETAEVIP
QESGRQTALFLKLASRWPIIHLHTDNGA"

```

```

gene
CDS

```

## ORIGIN

**RESULT 19**

**TITLE** Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology

**JOURNAL** J. Virol. 68 (11), 7433-7447 (1994)

**PUBMED** 7933127

**COMMENT** On Sep 9, 1994 this sequence version replaced gi:532151. Original source text: Human immunodeficiency virus type 2 (individual isolate FA) RNA.

**FEATURES**

source

Location/Qualifiers

1..708

/organism="Human immunodeficiency virus 2"

/mol\_type="genomic RNA"

/isolate="FA"

/db\_xref="taxon:11709"

/clone="1"

1..708

/gene="pol"

&lt;1..&gt;708

/gene="pol"

/codon\_start=2

/product="pol protein"

/protein\_id="AA56939.1"

/db\_xref="GI:533178"

/translation="MGIVVQPTSENRIVNIIEDMIKKEAIYVAVPAHKGIGNQ

EYDHLVSGIRQVFLERIEPAQEEHEKHSNIKELTHKFGIPQLVARQIVNTCAQCQ

OKGEATHGQVNAEIGVQMDCTHLEKIIIVAVHVASGFEAKVIPQSERQTALFL

KLASWPITLHLDNGSNFTSQEVKVAWVWVIGIEQTFGVYPNPQSQGVVAAWVHLLKN

QISRIREQANTTETIVL"

**ORIGIN**

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 13; Length 708;

Matches 19; Conservativity 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20

|||||

332 CAAGTAAATGCAGAAATAGG 351

**RESULT 21**

HIV2D194

LOCUS

DEFINITION

Human immunodeficiency virus type 2 (HIV-2), complete proviral

genome.

J04542

GI:325654

complete genome; core protein; env gene; gag gene; nef gene; pol

gene; regulatory factor; rev gene; reverse transcriptase; tat gene;

vif gene; vpr gene; vpx gene.

Human immunodeficiency virus 2 (HIV-2)

Human immunodeficiency virus 2

Viruses; Retro-transcribing viruses; Retroviridae;

Orthoretrovirinae; Lentivirus; Primate lentivirus group.

1 (bases 1 to 9472)

Kuehnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D.,

Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichener,C.,

Andreesen,R., Gelderblom,H. and Ruebsamen-Waigmann,H.

Unpublished

2 (sites)

Kuehnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D.,

Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichener,C.,

Andreesen,R., Gelderblom,H. and Ruebsamen-Waigmann,H.

Molecular cloning of two west African human immunodeficiency virus

type 2 isolates that replicate well in macrophages: a Gambian

isolate, from a patient with neurologic acquired immunodeficiency

syndrome, and a highly divergent Ghanian isolate

Proc. Natl. Acad. Sci. U.S.A. 86 (7), 2383-2387 (1989)

2467304

The revision in reference [2] corrected a possible frameshift

mutation in gag and a premature stop in env observed in the

original copy of the sequence.

[2] sites; for [1].

The gag coding sequence appears to be distorted at the 5' end by a

frameshift; the env cds has a premature stop codon. The 5' ends of

**FEATURES**

source

the pol and gag genes have not been determined.

Location/Qualifiers

1..9472

/organism="Human immunodeficiency virus 2"

/proviral

/mol\_type="genomic DNA"

/isolate="D194"

/db\_xref="taxon:11709"

1..299

/note="5"

/evidence=experimental

1..173

/standard\_name="R repeat"

/evidence=experimental

59..120

ctgggcagacg in [2] revises ag in [1]"

303..319

/bound\_moiety="Lys-tRNA primer"

/evidence=experimental

547..2112

/gene="gag"

547..2112

/gene="gag"

/codon\_start=1

/evidence=experimental

/protein\_id="AAA76840.1"

/db\_xref="GI:325657"

/translation="MGARNSVLRGKKADELKVLRPNPKRYLRKHVVVAAANELDRF

GLAESLLSKEGCKILKVLPLEVPTSGENKSLFNTVCVINCLEAEKVKDTTEAKK

LAQRHLVABTGAETKMPNISRPTAPSGKGNFPVQOAGNIIHVPLSEKVDTEAKK

VVEEKFGAEVWFGFQALSEGCTPYDINMLNCVGDHQAQWQIIRIINEEAAADWAQH

PFGPLPQALDRGSDIAGTTSTVDEQIOMYRQPNVPVGNIVYRWIIGLQKCV

RMYNTNILDVQKPKESFQSVYDRFYKSLRAEQTDPAVNMWMTLLIQQANPDCKL

VLKGLMNPTEMLTACQGVGSPQKARLMAEALKEALTAPIPFAAQORRAIRCW

NCKEGHSKACQCRAPRQCQKWKSGKSHIMANCPERQAGFLGPRGKQPRNPPAAQA

PQGLIPTAPPIDPAVDLLEKYQQQKQREQRPYKEVTEDLLHLEQETPRGATE

DILLHLSLFGKQD"

939..940

/gene="gag"

/note="ag in [2] revises agg in [1]"

&lt;1716..4937

/gene="pol"

&lt;1716..4937

/gene="pol"

/codon\_start=1

/evidence=experimental

/protein\_id="AAA76841.2"

/db\_xref="GI:7335662"

/translation="VLELWKGGLTGVSTPSTQKTGLLEVWQVTHHGKLPKGTGRFPFR

DGTGKAAPQLPRGSSSGADTNPNSSSGPGVEIYAAREKAEAEGETIQGGDGG

LTAPRAGRAPQGRGLATPOPSLWKPVTAFIEDQPVVELLDTGADDSIVAGIEL

GNYTPKIVGGIGGFINTKEYKNVEIKLVKVRATIMGDTPTINIFGRNIIATLWS

LNLPAKLDPIKVTLPKQDGRPKQWPLTKEIKELKEICEMEREGOLEEAPPTNP

YNTPTAIKKDKNKWRMLIDPRELNRVTQDFTETIQLGPHFAGLAKGKGLTVDVD

AVPSILHEDFROYTFTLPSVNNAPPEKRYVYKVLPGQWKGSPATFQVMMRQLLEPF

RKANPDVILIQYMDLILASDRTGLEHDKVLQLKELLNGLGFSFDEKFKQDPFQM

MGYELWPTWKQKIQIOLPOKEIWTVNDIQKLVGLNAAQIYPGIKTKHCLKIRGM

TLTREVQWTELAELAEENKILISQESGYQEEBELEATVIKSDQDQWAKIHOGE

RVLVKGYAKIKNTHNGVRLLAQVQKIGKEALVTNGRVKPHLPERVTWQWMDN

QYQVTVWPMDFVSTPLVRLTNLVGDDPIPGTETFTYDGSCKRQKGEKAGYVTDG

RDRVRVLEOTSNQAELEAFAMALADSPKPNVIVDSQYVMGIVAGOPTESNRINVQ

I1EDMTKBAVYVAVPAHKGIGNOBVDHLVSGQIRQVLFLEKIEPQBEHEKHYSN

IKELTHKFGIPQLVARQIVNTCAQCKQKEAIGHQVNAEIGVQWMDCTHLEKIIIVA

VHVASGFEAEVLPQSGRQTAFLKLASRWITLHLDNGGNFTSOEVKMAWMTIG

IEQSGFVYPNPQSGVYVAAWVHLLKNQISRIREQANTTETIIVLMAVCHMFKRGGIG

DMTPABRLINMITTEQIQLORKNNSNFKFYREGRDQLWKGPGELLWKGQGAVI

VKYGADIKVVPKRKAIIRDYGGRQELDSSSHLEGREDGEVA"

4867..5514

/gene="vif"

4867..5514

/gene="vif"

/codon\_start=1

/evidence=experimental  
/protein\_id="AAA76842.1"  
/db\_xref="GI:325659"  
/translation="MEEGKNWIVTWRVPGMRWHSLSVKHLKYRTKDLBEVRVYP  
HKGAWMTCSVIFPLEGESHLEIQAYWNLTPKGLWSSHSVRLTWYTKFPWTDVTP  
DCADSLIHSTFSCFAGEVRAIRGEKLLSCCNPQAKHQAQVPSLOYLALVVOQNG  
RPQRGAARKQWRDRHWRGLRVARQDYRSKQGGSEPSAPRAHPGVAKVLGILA"  
gene  
5342. .5680  
/gene="vpx"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76843.1"  
/db\_xref="GI:325660"  
/translation="WADPRVRVPPGNSGEETIGEAFWLDRDTIENLRVAVNHLPREL  
IFQWQRSWAYWHDSQGMSTSYRYLCIMQKAVYIHFKKGCTCLGRGHGPGGWRPG  
PPPPPPGLV"  
gene  
5680. .5997  
/gene="vpr"  
5680. .5997  
/gene="vpr"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76844.1"  
/db\_xref="GI:325661"  
/translation="WTEAPTEPPDDGTFRRELGSWTWVETLKEIKERALKHFDPCLL  
IALGNYIYNRHGDTLEGARELIRVLQALFVHIRAGCSRKGQTRRAPCPAAPTFR  
GMH"  
gene  
5843. .8371  
/gene="tat"  
/join(5843. .6138,8275. .8371)  
/gene="tat"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76845.1"  
/db\_xref="GI:325655"  
/translation="WETPLKEPESLESYNEPSSCTSRDVTVAQBRAXQGBELLAQLH  
RPLEATSKYQCSYHCQCLFKKGLGIWYARQGRRRRTPRKTKTHPPPSAKDSIS  
TRTGSQPTKQKQKTPETTVWSACGLH"  
gene  
6069. .8516  
/gene="rev1"  
/join(6069. .6138,8275. .8516)  
/gene="rev1"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76846.1"  
/db\_xref="GI:325656"  
/translation="WRDRADERGLOEKRLRLRLHQLINPYPHGQGTASORNNRRRRR  
RQWRLVALATKLHTIPDPPTDPLDRAIQQLGLTIQELPDPPDPLPESNSQGLAE  
T"  
gene  
6145. .8700  
/gene="env"  
6145. .8700  
/gene="env"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76847.1"  
/db\_xref="GI:325662"  
/translation="MEPGRNQLLVALLTSACLIYCKQVTVFYGIPAWRNASIPFLC  
ATKNRDWTGTOCLPNDYQBITLNVTEAPDANTVTEQAIEDVWRLFETSIKPCV  
KLTPUCVANNCTISGTTATPSPFNITIDENSTCIGNNCTGLGKEEVECEFNNTG  
LEQDKKRYNDWYSRDVCDDKNGTGYCMRHCNTSVIKESCDDHYDWMKFRYCAP  
PGFALLRNDNTSYGSEPKCSKVAASCTRMETQSTWFGNGTRAEANTYIYWHGK  
DNRTIISLKNYLNTHCKRPGNKTVPIITMSGRFRFSRPVYNNKKPGQAWCFQGNW  
TEAMREVKTAKHPRYGGTNDTKINFTKPGIGSDPEVTYMMWTCRGEFLYCNMTWF  
LNWVENKTNQTHGNAPCHIRQIINTWKKVGNVYLPREBELTCNSTVTSIINIDS  
DGNQNTIFSASVELYRLELDGYKLIEVPIPPAPTRKRYSSAPVRNKRGVFLVG  
LGLFATAGSAMGASLTLSAQRTLLAGIVOOQQLLDVVRKQBELRLTWGTKNLQ  
ARVTAIKYKDKQAGLNSWCFAFRQCHTTPWVNDSLTPDWNNTWQEWKRVHYLE  
ANISOLEAQIQQSKNMYLOKLSWDVFGNFDLTSWKYIQYGVTVVIGIIGIRI  
ATYIVQLLSRLKGRVPFSSPGVLYQIHIHTRGQPANRETEADGDSGFLGWPW  
PLNYIQFLIHLITRLITGLYNSCRGLLSKNSPTRLLISQSLTAIRDWLRLKAAVLYQG

variation  
6940. .6943  
/gene="rev1"  
/note="gaaa in [2] revises ga in [1]"  
6971. .6973  
/gene="rev1"  
/note="agg in [2] revises ag in [1]"  
6999. .7001  
/gene="rev1"  
/note="caa in [2] revises ca in [1]"  
7682. .7685  
/gene="rev1"  
/note="cgac in [2] revises cagc in [1]"  
7707. .7709  
/gene="rev1"  
/note="cgc in [2] revises cnc in [1]"  
8534. .9307  
/gene="nef"  
8534. .9307  
/gene="nef"  
/codon\_start=1  
Query Match 92.0%; Score 18.4; DB 13; Length 9472;  
Best Local Similarity 95.0%; Pred. NO. 4.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAATGCAGAAACAGG 20  
|||||  
Db 4212 CAAGTAATGCAGAAATAGG 4231  
RESULT 22  
HIVD194  
LOCUS HIVD194 9472 bp DNA linear VRL 18-APR-2005  
DEFINITION Human immunodeficiency virus type 2 strain D194 proviral genome.  
ACCESSION X52223  
VERSION X52223.1 GI:60155  
KEYWORDS env gene; envelope glycoprotein; gag gene; integrase; long terminal  
repeat; nef gene; pol gene; polymerase; reverse transcriptase;  
RNase H; tat gene; vif gene; vpr gene; vpx gene.  
SOURCE Human immunodeficiency virus 2 (HIV-2)  
ORGANISM Human immunodeficiency virus 2  
Viruses: Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.  
REFERENCE 1  
AUTHORS Kuehn,H., Kreutz,R. and Rubsamen-Waigmann,H.  
TITLE Nucleotide sequence of HIV-2D194, an isolate from a Gambian case of  
'neuro-AIDS', which showed excellent growth in macrophages  
JOURNAL Nucleic Acids Res. 18 (20), 6142 (1990)  
PUBMED 2235509  
REFERENCE 2 (bases 1 to 9472)  
AUTHORS Kuehn,H.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUN-1990) Kuehn,H., Chemotherapeutisches  
Forschungsinstitut, Georg-Speyer-Haus, Paul-Ehrlich-Strasse 42-44,  
600 Frankfurt 70, F R G  
COMMENT Virus DNA from a genomic library of infected cord blood  
lymphocytes. HIV-2 strain D194 was isolated in 1987 from  
peripheral blood of a Gambian male with neurological manifestations  
of AIDS.  
Ribosomal frame shift gives rise to pol gene product. See also  
<J04542> for overlapping sequences.  
FEATURES  
source  
1. .9472  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="D194"  
/db\_xref="taxon:11709"  
/cell\_type="cord blood lymphocyte"  
1. .173  
/note="R region"  
174. .299  
/note="U5 region"  
misc\_feature  
misc\_feature



KEYWORDS	Human immunodeficiency virus 2 (HIV-2)		
SOURCE	Human immunodeficiency virus 2		
ORGANISM	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.		
REFERENCE	1 (bases 1 to 9473)		
AUTHORS	Henco,K., von Briesen,H., Immelman,A., Kuehnel,H., Dietrich,U., Ruedsammen-waigmann,H. and Adamski,M.		
TITLE	HIV-2 virus variants		
JOURNAL	Patent: EP 0347365-A 1 20-DEC-1989;		
DIAGN	Institut fuer molekularbiologische Diagnostik GmbH;		
CHEMOTHERAPEUTISCHES FORSCHUNGSINSTITUT GEORG-SPEYER-HAUS			
FEATURES	Location/Qualifiers		
source	1..9473		
ORIGIN	/organism="Human immunodeficiency virus 2"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:11709"		
Query Match	92.0%; Score 18.4; DB 6; Length 9473;		
Best Local Similarity	95.0%; Pred. No. 4.3e+02;		
Matches	19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1 CAAGTAAATGCAGAACAGG 20	gene	
Db		CDS	
	4213 CAAGTAAATGCAGAAATAGG 4232		
RESULT 24			
HIV2U38293			
LOCUS	HIV2U38293 10312 bp DNA linear VRL 20-FEB-1997		
DEFINITION	Human immunodeficiency virus type 2, complete proviral genome.		
ACCESSION	U38293		
VERSION	U38293.1 GI:1845204		
KEYWORDS			
SOURCE	Human immunodeficiency virus 2 (HIV-2)		
ORGANISM	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.		
REFERENCE	1 (bases 1 to 10312)		
AUTHORS	Barnett,S.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-OCT-1995) Barton Bryant, HIV Database, Los Alamos National Laboratory, Eniwetok St., Los Alamos, NM 87545, USA		
Location/Qualifiers			
1..10312	/organism="Human immunodeficiency virus 2"		
/proviral			
/mol_type="genomic DNA"			
/db_xref="taxon:11709"			
/note="similar to HIV2BEN, GenBank Accession Number M30502, and to HIV2D194, GenBank Accession Number J04542"			
1103..2668	/gene="gag"	gene	
1103..2668	/gene="gag"	CDS	
/codon_start=1			
/evidence=not_experimental			
/product="gag polyprotein"			
/protein_id="AAB47780.1"			
/db_xref="GI:1845205"			
/translation="MGARNSVLRGKKADELEKIRLPGGKKYRLKHLIWAANEIDRFRLAENLSEKQCILVLDPLDPTSGENLSKLTVCVWCLHAEKVKDTEAAKLVQRHLVAETGAEMPTNSRTPAPSGKGNFPVQAGNYPVHPLSPRTINAWVKLVBEKFGAEVPGFQALSEGCTPYDINQMLNCVGGQAAQIIRIINEEAADAOHQVTPGPLAGQLDRPGSDIAGTSTVDQIOWMYQQNPVPGNIVRSWIOIGLOKCYRMYNFTNLDVKQPKESFSQSYDRFYSLSRAEQADPAVKNWMTOTLLVQANPCKLVKLGUMPTLEEMLTACGGVGGPKARLMAELKEALTTPPIPPAAQQRRTIRCWNCRGHSAKQCRACRQCGKCPGHIMANCPDQAGFLGMGRGKPRFPVAQAPQLIPTAPPADPAADLLEKYLQOQKORQREPRPYKEVTEDLLHLEQETPREATEDLLHLSLFGKQD"			
2863..5493		gene	
/gene="pol"		CDS	

2863..5493	/gene="pol"	CDS	
/codon_start=1			
/evidence=not_experimental			
/product="pol polyprotein"			
/protein_id="AAB47781.1"			
/db_xref="GI:1845206"			
/translation="MTGDTPIINFGRNLTALGMSLNPVARIPIKITLKPGDGKPRLKQWLTKEKVEALKEICERMEKEGQLEAPPTNPYPTPTFAIKKKQKRWMLIDFRNLNRPQDTYEQIQLGIPHPAGLAKKRITVLVDGDAYFSIDEDFRYTAFTLPSVNVNEVPGKYIVKVLPGQWKGSPAFQYTMROQLLEPERKANPDVILIQYMDLIIASDRTGLEHDKVLOLKELLNGLGFSTPEEKQKDPFQWGYELWPTWKLOKIQLOPKERTWTVDIOKLVLGILNAAQIYPGIKTKHLCLRLRGKMTLTEVQWTELAEAELEENRIILDQEGEHYQEEKLEATIQKSDNQWYTKHQEKLKVGKYAKIKNTHNTNGVRLLLAQVQKIGKEALVIGRWIKPHLPVERETEQMWDNQWVWIPWDFVSTPLVRLTFNLVGDPIPGTETFTYDGSQNRKSGKAGYVTDGRDKRVLEQTNQQAELAPAMALADSGPKVNIIVDSQYVGMIVASQPTSESIINQIIEDMIKKEAVYVAMTPAHKGIGNOBVDHLVSGIYQVLFLEKI BPAQBEHEKYSNIKELIHKFGI POLVARQIVNTCAOCQKGEAIIHQVNAEIGVMQDCTHLEGKIIIVAVHVASGFI EAEVTPQSGSGRTALFLKLASRWPIHTHTDNGPNTSQEVKMAVWVYIQSGFVGPYNPQSGQGVMEANHH LKNQISRIREQANTVETIVLMAVHCMPKRRGGIDMTFPAERLINMITTEQIEIQLQRKNSNFKFQVYVYREGRDQWKPGELLWKGEAVIVKVGADIKVVPKRKAKIIRDYGGROELDSSPHLEGAREDEGEVA"			
5423..6070	/gene="vif"	gene	
5423..6070	/gene="vif"	CDS	
/codon_start=1			
/evidence=not_experimental			
/product="vif protein"			
/protein_id="AAB47782.1"			
/db_xref="GI:1845207"			
/translation="MEEGKSWIVPTWRVPGRMERHSLVKYIKRTKDLGVRVYPHKVGMAMWTCRSRVIFPLKGESHLIEIQAYMNLTPKGLWSHSLVLTWYTKFPDVTDPDCADILIHSTYFCFAAGEVRRAIRGEKLSCCNYPQAHKAQVPSLOYLALVVQQND RPORKGPARKOWRRDRHWRGLVARQDYRSLERQCGKPPAPRAHPGVAKVLEILA"			
5898..6236	/gene="vpx"	gene	
5898..6236	/gene="vpx"	CDS	
/codon_start=1			
/evidence=not_experimental			
/product="vpx protein"			
/protein_id="AAB47783.1"			
/db_xref="GI:1845208"			
/translation="MTDPRVRVPPGNSGEETIGFAFWLDRITIALNREAVNHLPRELIFQVWQRSMRYWHDEQGLSSSYTKRYVLCMQKAMFIHLKKGCTCLGGGHPGGWRPGPPPPPPGLV"			
6341..8911	/gene="tat"	gene	
6341..8911	/gene="tat"	CDS	
/codon_start=1			
/product="tat protein"			
/protein_id="AAB47786.1"			
/db_xref="GI:1845211"			
/translation="METPLKAPESLSKNPSPCTSEWVAQAELAKQBELLAQLHLRPLEACTNPCKKCSFHCQLCKFLKGLGIWYVRKGGRRRTPKTKTHPPSTPPDKSISIQTQDSRPTKKQKCTPTPTVATTSLGR"			
6567..9056	/gene="rev"	gene	
6567..9056	/gene="rev"	CDS	
/codon_start=1			
/product="rev protein"			
/protein_id="AAB47787.1"			
/db_xref="GI:1845212"			
/translation="MCEKADSEELQRELRLRLHQTNPYPYRPGTAGQRNRRRRRRRQWLRLALADKLHTADPPPADSLDMAIQHQLQGLTIRELPDPTDLPESDSNQGIAET"			
6643..9240	/gene="env"	gene	
6643..9240	/gene="env"	CDS	





CDS	6239..6502 /note="vpr protein" /codon_start=1 /protein_id="AAB00740.1" /db_xref="GI:325651" /translation="WTEAPTFPPBDGTPRDLGSDWVIETLREIKEBALRHFDPRL IALGVYINRHGDTLEGARELIKTLQRALFVHFRAGCNRSIG" Join(6402..6697,8861..8957) /note="tat protein" /codon_start=1 /protein_id="AAB00741.1" /db_xref="GI:325646" /translation="METPLKAPESLSKYPNEPSSCTSERDVTQAQLAKOGHELLAQLH RPLEPCTNKCRCSCFQCQCFKGLGISTYRGRRRRPRKIKTSPSAPDKSIS TRTDSQPTKQKTSATVTVTCGLGQ" 6402..6697 /note="first expressed exon" /number=2 Join(6628..6697,8861..9102) /note="rev protein" /codon_start=1 /protein_id="AAB00742.1" /db_xref="GI:325647" /translation="MSERADEGLQGLRLRLRLHQTNPYQPGCTASQRNRRRRR RQWLRLVALANKLCAVPPTDPSLDRAIQLRLTIQELPDPPPTDLFESNQGLAE T" 6628..6697 /note="first expressed exon" /number=2 6704..9286 /note="env polyprotein" /codon_start=1 /protein_id="AAB00743.1" /db_xref="GI:325652" /translation="MEPCGNQLFVILLTSLACLVYCSQVTVFYGIPAWKNASIPLC ATKPRDTWGTQCLPDNDYQELILNVTAEADANNVTYEQAVEDWHLFETSIKPCV KLTPUCVAMNCSRGVQNTTTPNRTSSITSRPPTSAASINETSNCIENTCAGLG BEMMQCFNMGKQDKRRYKDTWYLEDVDCNDTGTCTYRHCNTSIIKESCDHY WDAMFRYCAPGFCALLKNDNTNYSGFPEKCTKVVAASCTRMETQTSTWFGNGRA ENRTYIYVHGRNRTIISLKYVNLTRCKRPNKTVLPITLMSGLVFHSQIPINTRPR QAWCFGRGWRAMQVKTLLVQHPRYKINDTGKINFTKPGAGSDPEVAFWNTNCRG EFLYCNMTWELNWDKQTRNRYCHIKQILNTHKVGKNVLPREGELACESTVTS LIANIDIKNRTHNITSAEVAELIRLELDYKLIETIPGPAPTDQRRVSSIFVRN KRGVFLGFLATAGSAMGARSITLSAQSRFLTAGIVQOOQLLDVVKQQEMRLR TWGNTQARVTAIEKYLKHOALNSWGCAPRQVCHTTPVWVNDLSLSPDMNMTWQE WEKVRYLEANIQSLEAQIQOENMYELQKLSNDILGNWFDLTSWVKYTIQYGVHI VVGIIALTAIVVOLLGRFKGRYPVFSPPGYLQOIHIHKDQOPANEGTEEDVGG DSGYLWMPNINYOFLIHLRLILGLYNICRDLKNSPTRLISLSLTAIRDMLR LKAALQVGCCEIQAFQAPARTTRETLAGAWGLWEAARRIGRILAVPRIRQAE LALL"
exon	8861..89102 /number=3 8861..8957 /number=3 9120..9893 /note="ref protein" /codon_start=1 /protein_id="AAB00744.1" /db_xref="GI:325653" /translation="WGASGSKLSKHSRCLRRLRARGDVGKORDASGGEYSOFQE ESGRQNSPSCGQQTQQGEYKNSFWNPATPRQKDLRYQNMDDVDSDDDLIGVPV TPRVPRRENTYKLAIDMSHFYKEKGLQGMFYRRRRHILDIYLEKEBGIIPDWQNYT HGPVRYPMYFGLWKLVSLEQAEDEANCLVHPAQTSRHHDEHGETTLVWQFDSM LAYNYKAPFLYPEEFCHKSGLPKEWKAKLKARGIPSE" 9505..10359 /note="3' LTR" 10211..10216
LTR	
polyA_signal	
ORIGIN	

Query Match 92.0%; Score 18.4; DB 13; Length 10359;  
Best Local Similarity 95.0%; Pred. No. 4.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	CAAGTAAATGCAGAAACAGG	20
Db	4768	CAAGTAAATGCAGAAATAGG	4787
RESULT	26		
LOCUS	AY341245		
DEFINITION	AY341245 Homo sapiens chromosome condensation 1-like (CHC1L) gene, complete cds.		
ACCESSION	AY341245		
VERSION	AY341245.1	GI:32891796	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 37676)		
AUTHORS	Rieder,M.J., Livingston,R.J., Daniele,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-JUL-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA		
COMMENT	To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).		
FEATURES	Location/Qualifiers		
source	1..37676	/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
misc_feature	1..346	/note="Region not scanned for variation"	
repeat_region	32..51	/rpt_family="L1"	
		/rpt_type="dispersed"	
repeat_region	119..405	/rpt_family="L1"	
		/rpt_type="dispersed"	
variation	557	/frequency="0.01"	
		/replace="c"	
gene	725..36573	/gene="CHC1L"	
mRNA	Join(725..833,3124..3219,3565..3629,9785..9940,10171..10321,12632..12798,13353..13511,13650..13757,14756..14898,22614..22804,23660..23786,25768..25907,29207..29337,35287..36573)	/gene="CHC1L"	
		/product="chromosome condensation 1-like"	
variation	852	/gene="CHC1L"	
		/frequency="0.01"	
		/replace="a"	
repeat_region	983..1159	/rpt_family="MIR"	
		/rpt_type="dispersed"	
variation	1085	/gene="CHC1L"	
		/frequency="0.01"	
		/replace="g"	
variation	1096	/gene="CHC1L"	
		/frequency="0.01"	
		/replace="c"	
variation	1327	/gene="CHC1L"	
		/frequency="0.04"	
		/replace="c"	
variation	1383	/gene="CHC1L"	



WPCOMMENT

Sequence split into 23 fragments LOCUS CR382125 Accession CR382125

Fragment Name	Begin	End
CR382125_00	1	110000
CR382125_01	100001	210000
CR382125_02	200001	310000
CR382125_03	300001	410000
CR382125_04	400001	510000
CR382125_05	500001	610000
CR382125_06	600001	710000
CR382125_07	700001	810000
CR382125_08	800001	910000
CR382125_09	900001	1010000
CR382125_10	1000001	1110000
CR382125_11	1100001	1210000
CR382125_12	1200001	1310000
CR382125_13	1300001	1410000
CR382125_14	1400001	1510000
CR382125_15	1500001	1610000
CR382125_16	1600001	1710000
CR382125_17	1700001	1810000
CR382125_18	1800001	1910000
CR382125_19	1900001	2010000
CR382125_20	2000001	2110000
CR382125_21	2100001	2210000
CR382125_22	2200001	2234072

Continuation (5 of 23) of CR382125 from base 400001 (CR382125 Kluyveromyces lactis strain

Query Match 90.0%; Score 18; DB 15; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19

Db 69477 AAGTAAATGCAGAAACAG 69494

RESULT 28

AL157813/C

LOCUS

DEFINITION

AL157813

Human DNA sequence from clone Rp11-165D7 on chromosome 13 Contains

the CHC1L gene for chromosome condensation 1-like, 3 novel genes

and a CpG island, complete sequence.

AL157813.11 Gi:13160187

HTG; CHC1L; CpG island.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 114519)

AUTHORS Oliver.K.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Feb 27, 2001 this sequence version replaced gi:12743755.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

Rp11-165D7 is from the library RPC1-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)

-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

source

1. 114519  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-165D7"  
/clone\_lib="RPC1-11.1"  
1  
/note="Clone left end: RP11-165D7"  
complement(join(554..1842,7782..7912,11212..11351,  
13333..13459,14315..14505,22220..22362,23361..23468,  
23607..23765,24320..24486,26796..26946,27177..27332,  
33489..33553,33899..33994,36250..36348,44619..44751))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(join(554..1842,7782..7912,11212..11351,  
13333..13459,14315..14505,22220..22362,23361..23468,  
23607..23765,24320..24486,26796..26946,27177..27332,  
33489..33553,33899..33994,36250..36348,44619..44751))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
/product="chromosome condensation 1-like"  
/note="match: ESTs: AA116579 AI245932 AL596966 AL702635  
AU129309 AU132093 AU138162 AV711315 AW210388 BB851525  
BF287003 BF416274 BF901505 BG027293 BI667346 BI756233  
BI918241 BI918675 BM227327 BM228676 BM681889 BM723702  
BQ184658  
match: CDNAS: AF060219 AK023010 BC029052"  
complement(554)  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(577..582)  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(join(1488..1842,3243..3306,4031..4164))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-003"  
complement(join(1488..1842,3243..3306,4031..4164))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-003"  
/product="chromosome condensation 1-like"  
/note="match: ESTs: BI438387"  
complement(join(1702..1842,7782..7912,11212..11351,  
13333..13459,14315..14505,22220..22362,23361..23468,  
23607..23765,24320..24486,26796..26946,27177..27332,  
33489..33530))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
/standard\_name="OTTHUMP00000018399"  
/note="match: proteins: O95199 Q99LJ7 Q9NV13"  
/codon\_start=1  
/product="chromosome condensation 1-like"  
/protein\_id="CAH70795.1"  
/db\_xref="GI:55661682"  
/db\_xref="GeneW:1914"  
/db\_xref="GOA:O95199"  
/db\_xref="InterPro:IPR000210"  
/db\_xref="UniProt/TREMBL:O95199"  
/translation="MEBELPLFGSDSGKPVQATLSSKMLDVGKWPFLSCSEELQL  
IRQACVFGSAGNEVLYTVNDEIFVLGTNCCGLGQVQSTIEPRDLSLNGKIAIC

misc\_feature

gene

mRNA

polyA\_site

polyA\_signal

gene

mRNA

CDS

```

LSYSGPHIVLATTEGEVFWGNAYSQLNGTTHNLGLVPCHEISTNLSNKQVIEVACG
SVHSLVLTGDEGVANGNNGVSGSTVNOPIPRRTVTCGLQNKVGVYATACGCMCCM
AVUDTGVTVGVNGQGLGNSGNQTPCVAALQGRVORVACGYAHTLVLTDEG
QYVANGANSYQGLGTGNKSNQSPYPTVEKDRILEIAACHSTHTSAKTQCGHYVM
GQCGQSVILPHLTFTSCDDDFACFPATVWRLLSVDPDDHLTVAESLSKREFDNP
TADLFLVQGRYIAHKVLLKRCFATRESLEDNEDDIVESEFSYPVYRAFLVLYDT
DSISPEBAVGLLDIATYRNLKLCQQTIKQICENAIALLSAVKRYDAQDLE
EFCFRFCINHLTVVTQTSFAEMDHLKFNFTSKASRVGAFKN"
3512
/note="Clone_right_end: RP11-17A110"
14623..14650
/note="Single clone region. Assembly confirmed by
restriction digest data."
/gene tag="RP11-165D7.2-001"
complement(16657..18528)
/mRNA tag="RP11-165D7.2-001"
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
complement(16657..18528)
/standard_name="OTTHUMP0000018400"
/note="match: proteins: Q96DJ9"
/codon_start=1
/protein_id="CAH70794.1"
/db_xref="GI:55661681"
/db_xref="UniProt/TREMBL:OSVZX9"
/translation="MALQLRWFKHRRPFFILLIAQLCVHWPHELPPDQRFCLPOHCO
PGVFLDSSGPPSPAPPSALTSPRSLLGLNLAWGLSQSRFLFSLAACLAAHC
GOVLSSENTISHHGVKAVSLHPLTSDOIRTLRWIEADATVPKRPPIPALAS
VPRVSTMLLSAILPSPALPACWCSIALHTDSYSTALASPKHTWPS"
complement(join(23723..23765,24320..24486,26796..26946,
27177..27332,44619..44771))
/gene="CHC1L"
/locus tag="RP11-165D7.1-002"
complement(join(23723..23765,24320..24486,26796..26946,
27177..27332,44619..44771))
/gene="CHC1L"
/locus tag="RP11-165D7.1-002"
/product="chromosome condensation 1-like"
/note="match: ESTs: BI754136 BI754283"
complement(join(82230..82432,84699..84908))
/locus tag="RP11-165D7.3-001"
complement(join(82230..82432,84699..84908))
/locus tag="RP11-165D7.3-001"
complement(join(82230..82432,84699..84908))
/note="match: ESTs: AI769787"
complement(join(88565..89450,89604..89679))
/locus tag="RP11-165D7.4-001"
complement(join(88565..89450,89604..89679))
/locus tag="RP11-165D7.4-001"
complement(join(88565..89450,89604..89679))
/note="match: ESTs: AI307356 AI762913 AI767356 AI796352
AW469968 BM472085"
114420
/note="Clone_left_end: RP11-108P5"

ORIGIN
Query Match 90.0%; Score 18; DB 8; Length 114519;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAACAG 19
|||||
Db 30030 AAGTAAATGCAGAACAG 30013

RESULT 29
AL833774/c
LOCUS AL833774 194045 bp DNA linear ROD 09-AUG-2002
DEFINITION Mouse DNA sequence from clone RP23-383116 on chromosome 2, complete
sequence.
ACCESSION AL833774
VERSION AL833774.4 GI:22204783
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)

```

## ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

## Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 194045)

## Tracey, A.

## Direct Submission

Submitted (06-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 11, 2002 this sequence version replaced gi:21953137.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-383116 is  
from the RP23-23 Mouse PAC Library

constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6.

## FEATURES

## Source

Location/Qualifiers  
1..194045  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone\_lib="RP23-383116"  
/clone\_lib="RP23-23"

## ORIGIN

Query Match 90.0%; Score 18; DB 9; Length 194045;  
Best Local Similarity 100.0%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 1 CAAGTAAATGCAGAACACA 18

|||||

## Db 94793 CAAGTAAATGCAGAACACA 94776

RESULT 30  
AC134627  
LOCUS

DEFINITION AC134627 204712 bp DNA linear HTG 20-NOV-2002  
Rattus norvegicus clone CH230-368N4, WORKING DRAFT SEQUENCE, 5  
unordered pieces.

## ACCESSION

AC134627.2 GI:25138679

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

## SOURCE

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

## AUTHORS

1 (bases 1 to 204712)  
Muzny, D.Marie., Metzker, M.Lee., Abramzon, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazoe, I., Ceasear, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Labow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulaeged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: KC20  
Center clone name: CH230-368N4  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 186320 bases at least Q40  
Consensus quality: 188335 bases at least Q30  
Consensus quality: 189791 bases at least Q20  
Estimated insert size: 187550; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES	Location/Qualifiers
1	6702: contig of 6702 bp in length
6703	6802: gap of unknown length
6803	76788: contig of 69986 bp in length
76789	76888: gap of unknown length
76889	127999: contig of 51011 bp in length
127900	127999: gap of unknown length
128000	153581: contig of 25582 bp in length
153582	153681: gap of unknown length
153682	204712: contig of 51031 bp in length.
source	1..204712
	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-368N4"
gap	6703..6802
misc_feature	/estimated_length=unknown
	8626..9525
	/note="clone boundary"
	clone_end:Spf
	site:
	end_sequence:BZ168747"
gap	76789..76888
misc_feature	/estimated_length=unknown
	76889..78500
	/note="wgs_contig"
gap	127900..127999
	/estimated_length=unknown
gap	153582..153681
	/estimated_length=unknown
misc_feature	177120..178477
	/note="wgs_contig"
ORIGIN	
Query Match	90.0%; Score 18; DB 14; Length 204712;
Best Local Similarity	100.0%; Pred. No. 6e+02;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	3 AGTAATGACAGG 20
Db	134800 AGTAATGACAGG 134817
RESULT 31	
AC095481	

On Nov 20, 2002 this sequence version replaced gi:23343609.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome

LOCUS	AC095481	276881 bp	DNA	linear	HTG 09-MAY-2003			
DEFINITION	Rattus norvegicus clone CH230-7K20, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.							
ACCESSION	AC095481							
VERSION	AC095481.6	GI:30467658						
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.							
SOURCE	Rattus norvegicus (Norway rat)							
ORGANISM	Rattus norvegicus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.							
AUTHORS	1 (bases 1 to 276881)							
	Muzny, D. Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Kocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanea, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.							
	Direct Submission							
	Unpublished							
REFERENCE	2 (bases 1 to 276881)							
AUTHORS	Worley, K. C.							
TITLE	Direct Submission							
JOURNAL	Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA							
REFERENCE	3 (bases 1 to 276881)							
AUTHORS	Rat Genome Sequencing Consortium.							
TITLE	Direct Submission							
JOURNAL	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department							

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On May 9, 2003 this sequence version replaced gi:24941064.	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
----- Genome Center	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
----- Project Information	
Center project name: GCKW	
Center clone name: CH230-7K20	
----- Summary Statistics	
Assembly program: Atlas;	
Consensus quality: 22954 bases at least Q40	
Consensus quality: 228004 bases at least Q30	
Consensus quality: 230763 bases at least Q20	
Estimated insert size: 23155; sum-of-contigs estimation	
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation	
-----	
* NOTE: Estimated insert size may differ from sequence length	
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)	
* NOTE: This is a "working draft" sequence. It currently	
* consists of 5 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 241308: contig of 241308 bp in length	
* 241309 241408: gap of unknown length	
* 241409 270462: contig of 23054 bp in length	
* 270463 270562: gap of unknown length	
* 270563 271681: contig of 1119 bp in length	
* 271682 271781: gap of unknown length	
* 271782 272810: contig of 1029 bp in length	
* 272811 272910: gap of unknown length	
* 272911 276881: contig of 3971 bp in length.	
Location/Qualifiers	
1. .276881	
/organism="Rattus norvegicus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10116"	
/clone="CH230-7K20"	
1. .1235	
/note="wgs_contig"	
63196. .65045	
/note="wgs_contig"	
167109. .158091	
/note="wgs_contig"	
241309. .241408	
/estimated_length=unknown	
270463. .270562	
/estimated_length=unknown	
271682. .271781	
/estimated_length=unknown	
272811. .272910	
/estimated_length=unknown	
ORIGIN	
Query Match	90.0%; Score 18; DB 14; Length 276881;

## COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 9, 2003 this sequence version replaced gi:24941064.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GCKW  
 Center clone name: CH230-7K20  
 ----- Summary Statistics  
 Assembly program: Atlas;  
 Consensus quality: 222954 bases at least Q40  
 Consensus quality: 228004 bases at least Q30  
 Consensus quality: 230763 bases at least Q20  
 Estimated insert size: 233155; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 241308: contig of 241308 bp in length  
 \* 241309 241408: gap of unknown length  
 \* 241409 270462: contig of 29054 bp in length  
 \* 270463 270562: gap of unknown length  
 \* 270563 271681: contig of 1119 bp in length  
 \* 271682 271781: gap of unknown length  
 \* 271782 272810: contig of 1029 bp in length  
 \* 272811 276881: gap of unknown length  
 \* 272911 276881: contig of 3971 bp in length.

## FEATURES

source  
 1. 276881  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-7K20"  
 misc\_feature  
 1. 1235  
 /note="wgs contig"  
 misc\_feature  
 63196. 65045  
 /note="wgs contig"  
 misc\_feature  
 167109. 168091  
 /note="wgs contig"  
 gap  
 241309. 241408  
 /estimated\_length=unknown  
 gap  
 270463. 270562  
 /estimated\_length=unknown  
 gap  
 271682. 271781  
 /estimated\_length=unknown  
 gap  
 272811. 272910  
 /estimated\_length=unknown

## ORIGIN

Query Match 90.0%; Score 18; DB 14; Length 276881;

Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAATAATGCAGAAACAGG 20  
|||||  
Db 247941 AGTAATAATGCAGAAACAGG 247958

RESULT 32  
G52355/c  
LOCUS SHGC-85070 Human Homo sapiens STS genomic, linear STS 30-MAR-2000  
DEFINITION  
ACCESSION G52355  
VERSION G52355.1 GI:5223682  
KEYWORDS STS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Olivier, M. and Cox, D.R.  
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)  
JOURNAL Unpublished (2000)  
COMMENT

Contact: Michael Olivier, David R. Cox  
Stanford Human Genome Center  
Stanford University School of Medicine  
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
Tel: (650) 320-5800  
Fax: (650) 320-5801  
Email: olivier@shgc.stanford.edu  
Primer A: GGAGAACTCGGTGAGAGAAATT  
Primer B: GTGCTAAATGACACCCACATGAA  
STS size: 286

PCR Profile:  
Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Amplifitq Gold Polymerase: 0.07 units/ul  
Total Vol: 5 ul

Buffer:  
MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES  
source  
1..491  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="12"  
/clone\_lib="Human"  
191..476  
191..213  
complement(454..476)  
STS  
primer\_bind  
primer\_bind  
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 491;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19  
|||||  
Db 242 CAAGTAATGCAGAAACAG 224

RESULT 33  
G90172/c  
LOCUS S210P603RB3.T0 BALB/cByJ Mus musculus STS genomic, linear STS 06-SEP-2002  
DEFINITION  
ACCESSION G90172  
VERSION G90172.1 GI:22740928  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Wade, C.  
TITLE Polymorphism Structure in the Mouse  
JOURNAL Unpublished (2002)  
COMMENT

Contact: Kerstin Lindblad-Toh  
Whitehead Institute for Biomedical Research, Center for Genome Research  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172521477  
Fax: 6172580903  
Email: kersli@genome.wi.mit.edu  
Primer A: None  
Primer B: None  
STS size: 549

Protocol:  
WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES  
Location/Qualifiers  
1..549  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="BALB/cByJ"  
/db\_xref="taxon:10090"  
/map="- 1 22-488 157394789-157395256"  
/clone\_lib="BALB/cByJ"  
<1..>549

STS  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 10; Length 549;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20  
|||||  
Db 322 AAGTAATGCAGAAACAGG 304

RESULT 34  
BV558070  
LOCUS gmm55f08.g1 Clint Pan troglodytes versus STS genomic, linear STS 09-APR-2005  
DEFINITION  
ACCESSION BV558070  
VERSION BV558070.1 GI:62449091  
KEYWORDS STS.  
SOURCE Pan troglodytes versus  
ORGANISM Pan troglodytes versus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan.  
 1 (bases 1 to 723)  
 Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
 Jaffe,D.B.  
 Initial Sequence of the Chimpanzee Genome and Comparison with the  
 Human Genome  
 Unpublished (2005)  
 Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580903  
 Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 723  
 Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan  
 troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald, Karlén, Yvonne), 3 Pan  
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps  
 of unknown origin  
 (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan  
 troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred  
 score >= 20, at least 30% of its base calls must satisfy  
 SNQS(30,25) (single strand NQS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred  
 score >= 25), and the read must have at least 200 bp SNQS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not  
 consistently placed were  
 discarded. After above filtering, NQS(30,25) standard was applied  
 to all pairs of  
 overlapping reads to call NQS bases and SNPs. Alignments (between  
 two reads) with less  
 than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
 exclude alignment between two  
 copies of a single read,  
 95% of their genome  
 alignments (>=95% bases of read A and >=95% bases of read B were  
 placed at the same locus  
 of human genome) were discarded.  
 Location/Qualifiers  
 1..723  
 /organism="Pan troglodytes verus"  
 /mol\_type="genomic DNA"  
 /sub\_species="verus"  
 /db\_xref="taxon:37012"  
 /clone\_lib="Clint"  
 <1..>723

STSSource

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 723;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGCTAATGCAGAACAGG 20  
 |||||

Db 641 AAGTAAATGCAGAACAGG 659  
 |||||

RESULT 35  
 BV586373/c

LOCUS BV586373 737 bp DNA linear STS 12-APR-2005  
 DEFINITION G591P63187RH10.T0 Clint Pan troglodytes verus STS genomic,  
 sequence tagged site.  
 ACCESSION BV586373  
 VERSION BV586373.1 GI:62502102  
 KEYWORDS STS.  
 SOURCE Pan troglodytes verus  
 ORGANISM Pan troglodytes verus  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan.  
 1 (bases 1 to 737)  
 Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
 Jaffe,D.B.  
 Initial Sequence of the Chimpanzee Genome and Comparison with the  
 Human Genome  
 Unpublished (2005)  
 Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580903  
 Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 737  
 Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan  
 troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald, Karlén, Yvonne), 3 Pan  
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps  
 of unknown origin  
 (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan  
 troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred  
 score >= 20, at least 30% of its base calls must satisfy  
 SNQS(30,25) (single strand NQS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred  
 score >= 25), and the read must have at least 200 bp SNQS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not  
 consistently placed were  
 discarded. After above filtering, NQS(30,25) standard was applied  
 to all pairs of  
 overlapping reads to call NQS bases and SNPs. Alignments (between  
 two reads) with less  
 than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
 exclude alignment between two  
 copies of a single read,  
 95% of their genome  
 alignments (>=95% bases of read A and >=95% bases of read B were  
 placed at the same locus  
 of human genome) were discarded.  
 Location/Qualifiers  
 1..737  
 /organism="Pan troglodytes verus"  
 /mol\_type="genomic DNA"  
 /sub\_species="verus"  
 /db\_xref="taxon:37012"  
 /clone\_lib="Clint"  
 <1..>737

STSSource

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 737;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAACAGG 20

Db 367 AAGTAAATGCAGAACAGG 349

RESULT 36

BV576492

LOCUS

DEFINITION BV576492 800 bp DNA linear STS 12-APR-2005

tagged site.

ACCESSION BV576492

VERSION BV576492.1 GI:62492221

KEYWORDS STS.

SOURCE Pan troglodytes verus

ORGANISM Pan troglodytes verus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

1 (bases 1 to 800)

Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and

Jaffe,D.B.

Initial Sequence of the Chimpanzee Genome and Comparison with the

Human Genome

Unpublished (2005)

CONTACT: Michael C. Zody

Broad Institute of MIT and Harvard

320 Charles Street, Cambridge, MA 02141, USA

Tel: 6172580933

Fax: 6172580903

Email: mczody@broad.mit.edu

Primer A: No sequence submitted

Primer B: No sequence submitted

STS size: 800

Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin

(Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of

overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less

than 100 NQS bases or with SNP rate > 0.01 were discarded. To

exclude alignment between two

copies of a single read, comparisons between two reads that share

95% of their genome

alignments (>=95% bases of read A and >=95% bases of read B were

placed at the same locus

of human genome) were discarded.

Location/Qualifiers

1. .800

/organism="Pan troglodytes verus"

/mol\_type="genomic DNA"

/sub\_species="verus"

/db\_xref="taxon:37012"

/clone\_lib="Clint"

<1..>800

STS

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 800;

Best Local Similarity 94.7%; Pred. No. 1.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAACAGG 20

Db 698 AAGTAAATGCAGAACAGG 716

RESULT 37

BV628207

LOCUS

DEFINITION BV628207 800 bp DNA linear STS 15-APR-2005

Sequence tagged site.

ACCESSION BV628207

VERSION BV628207.1 GI:62618385

KEYWORDS STS.

SOURCE

ORGANISM

Pan troglodytes troglodytes

Pan troglodytes troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

1 (bases 1 to 800)

Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and

Jaffe,D.B.

Initial Sequence of the Chimpanzee Genome and Comparison with the

Human Genome

Unpublished (2005)

JOURNAL

COMMENT

Contact: Michael C. Zody

Broad Institute of MIT and Harvard

320 Charles Street, Cambridge, MA 02141, USA

Tel: 6172580933

Fax: 6172580903

Email: mczody@broad.mit.edu

Primer A: No sequence submitted

Primer B: No sequence submitted

STS size: 800

Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI

Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,

including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps

(Donald, Karlén, Yvonne), 3 Pan

troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps

of unknown origin

(Gon, Unknown Chimp). Common names: Pan troglodytes verus is the

western chimp and Pan

troglodytes troglodytes is the central chimp. To be included in

chimpanzee SNP discovery, a

read must be at least 500bp in length, at least 50% of its base

calls must have Phred

score >= 20, at least 30% of its base calls must satisfy

SNQS(30,25) (single strand NQS, the

base in question has Phred score >= 30, the surrounding 10 bases in

the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25)

bases. Reads not uniquely

placed in the genome and read pairs whose two ends were not

consistently placed were

discarded. After above filtering, NQS(30,25) standard was applied

to all pairs of

overlapping reads to call NQS bases and SNPs. Alignments (between

two reads) with less

than 100 NQS bases or with SNP rate > 0.01 were discarded. To

exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome  
alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus  
of human genome) were discarded.

## FEATURES

source  
1. .800  
/organism="Pan troglodytes troglodytes"  
/mol\_type="genomic DNA"  
/sub\_species="troglodytes"  
/db\_xref="taxon:37011"  
/clone\_lib="Noemie"  
<1. .>800

## STS

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 800;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20

DB 412 AAGTAATGCAGAAACAGG 430

## RESULT 38

BV594335/c

LOCUS BV594335 806 bp DNA linear STS 12-APR-2005  
DEFINITION taw08f01.b1 Clint Pan troglodytes versus STS genomic, sequence tagged site.

ACCESSION BV594335

VERSION BV594335.1 GI:62510064

KEYWORDS STS.

SOURCE Pan troglodytes versus

ORGANISM Pan troglodytes versus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eumetazoa; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.

1 (bases 1 to 806)

REFERENCE Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and

Jaffe,D.B.

Initial Sequence of the Chimpanzee Genome and Comparison with the

Human Genome

Unpublished (2005)

## JOURNAL

## COMMENT

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA

Tel: 6172580933

Fax: 6172580903

Email: mczody@broad.mit.edu

Primer A: No sequence submitted

Primer B: No sequence submitted

STS size: 806

Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan

troglodytes versus), 3 other Pan

troglodytes (Noemie, Masuku, Clara) and 2 chimps

troglodytes troglodytes chimps

of unknown origin

(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the

western chimp and Pan

troglodytes troglodytes is the central chimp. To be included in

chimpanzee SNP discovery, a

read must be at least 500bp in length, at least 50% of its base

calls must have Phred

score >= 20, at least 30% of its base calls must satisfy

SNQS(30,25)(single strand NQS, the

base in question has Phred score >= 30, the surrounding 10 bases in

the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

## source

1. .806  
/organism="Pan troglodytes versus"  
/mol\_type="genomic DNA"  
/sub\_species="versus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1. .>806

## STS

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 806;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20

DB 549 AAGTAATGCAGAAACAGG 531

## RESULT 39

BC063923/c

LOCUS BC063923

DEFINITION Xenopus tropicalis hypothetical protein MGC76215, mRNA

2081 bp linear VRT 23-AUG-2004

MGC:76215 IMAGE:5336233, complete cds.

ACCESSION BC063923

VERSION BC063923.1

KEYWORDS GI:39795467

SOURCE MGC.

ORGANISM

Xenopus tropicalis

(Silurana tropicalis)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 2081)

REFERENCE Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heideh,F.,

Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,

Stapleton,M., Soares,M.B., Donaldson,M.F., Casavant,T.L.,

Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,

Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,

McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,

Souffard,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,

Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,

Schneerch,A., Schein,J.E., Jones,S.J., Skalska,U., Smalios,D.E.,

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2081)

REFERENCE Klein,S. and Gerhard,D.S.

AUTHORS

TITLE

Direct Submission

JOURNAL	Submitted (08-DEC-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA			REFERENCE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
REMARK	Contact: XGC help desk			AUTHORS	Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayaashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Iehi, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Negahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hoshi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Kamiyama, K., Katsuta, N., Sato, K., Kikkawa, E., Omura, Y., Abe, K., Nishikawa, Y., Matsunawa, H., Ichihara, T., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoteuchi, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togai, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Muesashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.		
source	Complete sequencing and characterization of 21,243 full-length human cDNAs			TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs		
gene	1. .2081			JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)		
CDS	/organism="Xenopus tropicalis"			PUBMED	14702039		
	/mol_type="mRNA"			REFERENCE	Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Iehi, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.		
	/db_xref="GeneID:394910"			AUTHORS	Isogai, T. and Yamamoto, J.		
	/db_xref="GeneID:394910"			TITLE	Unpublished		
	/codon_start=1			JOURNAL	3 (bases 1 to 2286)		
	/product="hypothetical protein MGC76215"			REFERENCE	Isogai, T. and Yamamoto, J.		
	/protein_id="AAH63923.1"			AUTHORS	Direct Submission		
	/db_xref="GI:39795468"			TITLE	Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan		
ORIGIN	/translation="MEALFIFSLIDCCALIFLSVVFITLSDLECDYINARSCCKL NKWVPELVGHTVSVLMLVSLHWFILNLPVAAWNIYRFIMVPSGNLGVDPDTEIH NRGQLKSHMKEAMIKLGFHLLCFPIYLYSMILALIND"			JOURNAL	(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.		
Query Match	87.0%;	Score 17.4;	DB 5; Length 2081;	COMMENT	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.		
Best Local Similarity	94.7%;	Pred. No. 1.4e+03;			Location/Qualifiers		
Matches	18;	Conservative 0;	Mismatches 1; Indels 0; Gaps 0;		1. .2286		
Qy	1	CAAGTAAATCGAAGAACAG 19			/organism="Homo sapiens"		
Db	409	CAAGTAAATCGAAGAACAG 391			/mol_type="mRNA"		
RESULT 40					/db_xref="taxon:9606"		
AK123724/c					/clone="HLUNG2017262"		
LOCUS	AK123724	2286 bp	mRNA linear		/tissue_type="lung"		
DEFINITION	Homo sapiens cDNA FLJ41730 fis, clone HLUNG2017262.				/clone_lib="HLUNG2"		
ACCESSION	AK123724				/note="cloning vector: pME18SFL3"		
VERSION	AK123724.1			ORIGIN			
KEYWORDS	oligo capping; fis (full insert sequence).			Query Match	87.0%;	Score 17.4;	DB 8; Length 2286;
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						





```
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAACAGG 20
Db 1405 AGGTAATGCAGAACAGG 1387

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 3192: contig of 3192 bp in length
* 3292: gap of 100 bp
* 3293: contig of 4453 bp in length
* 7745: gap of 100 bp
* 7845: gap of 100 bp
* 14184: contig of 6339 bp in length
* 14284: gap of 100 bp
* 14285: contig of 2576 bp in length
* 14285: contig of 2576 bp in length
* 16960: gap of 100 bp
* 16961: gap of 100 bp
* 18961: contig of 2350 bp in length
* 19310: contig of 100 bp
* 19311: contig of 3360 bp in length
* 19410: gap of 100 bp
* 22770: gap of 100 bp
* 22771: gap of 100 bp
* 22871: contig of 4233 bp in length.

FEATURES             Location/Qualifiers
    source             1..27103
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /clone="RP3-432N11"
                        /clone_lib="RPCI-3"
    misc_feature        1..3192
                        /note="assembly_fragment:00004"
    misc_feature        3293..7745
                        /note="assembly_fragment:00264"
    misc_feature        7846..14184
                        /note="assembly_fragment:00291"
    misc_feature        14285..16860
                        /note="assembly_fragment:00398"
    misc_feature        16961..19310
                        /note="assembly_fragment:00512"
    misc_feature        19411..22770
                        /note="assembly_fragment:00623"
    misc_feature        22871..27103
                        /note="assembly_fragment:00707"

ORIGIN
Query Match          87.0%; Score 17.4; DB 14; Length 27103;
Best Local Similarity 94.7%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAACAGG 20
Db 10933 AAGTATATGCAGAACAGG 10915

-----
RESULT 48
LOCUS HSE129H9 37170 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone JL22NC01-129H9 on chromosome 22,
complete sequence.
ACCESSION 268224
VERSION 268224.1 GI:1122885
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAACAGG 20
Db 1405 AGGTAATGCAGAACAGG 1387

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 3192: contig of 3192 bp in length
* 3292: gap of 100 bp
* 3293: contig of 4453 bp in length
* 7745: gap of 100 bp
* 7845: gap of 100 bp
* 14184: contig of 6339 bp in length
* 14284: gap of 100 bp
* 14285: contig of 2576 bp in length
* 14285: contig of 2576 bp in length
* 16960: gap of 100 bp
* 16961: gap of 100 bp
* 18961: contig of 2350 bp in length
* 19310: contig of 100 bp
* 19311: contig of 3360 bp in length
* 19410: gap of 100 bp
* 22770: gap of 100 bp
* 22771: gap of 100 bp
* 22871: contig of 4233 bp in length.

FEATURES             Location/Qualifiers
    source             1..7886
                        /organism="Drosophila sp."
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:7242"

ORIGIN
Query Match          87.0%; Score 17.4; DB 6; Length 7886;
Best Local Similarity 94.7%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAACAGG 20
Db 3324 AGGTAATGCAGAACAGG 3306

-----
RESULT 47
LOCUS AL353656/2 27103 bp DNA linear HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 6 clone RP3-432N11, 7 unordered pieces.
ACCESSION AL353656
VERSION AL353656.2 GI:9863650
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.

REFERENCE
1 Sims.S.
Direct Submission
Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9213126.

-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: dJ432N11
-----
Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 22990 bases at least Q40
```





gene  
KVNHNLSGVIPOGRQNTFDAESYFGNRLLCQPTNRSNNNSYEADNGVEADESI  
IDMVSFLGFAAAVYVILIGILASLDFSPWRFYKYVDAFIKVRNLLL"  
/gene="F9E11.6"  
/notes="similar to GI:1184077 from [Lycopodium  
pimpinellifolium]"  
join(<6346..6733,7017..7172,7273..7341,7661..7735,  
7890..>10057)  
/gene="F9E11.6"  
/join(6346..6733,7017..7172,7273..7341,7661..7735,  
7890..10057)  
/codon start=1  
/product="disease resistance protein, putative;  
6346-10057"  
/protein\_id="AAG51872.1"  
/db\_xref="GI:12323813"  
/translation="MERKPSGQQLIWMVLLLVOLRGYKCCIEKERKALLEKKYMS  
KPADGLSDVLPWTNDTKSNCCRWEGLKCNQTSRIIEISIQTNFKESLLNLSLL  
HPEELRSNLNGEINYEENGLFDDVEGESLRRLRNLEITLDSNSFNNSIPFVLNA  
ATSLTLFTQSNVIGGLPIKELKNLTKLELDSRGYNGSIPELKVLTNLVLGLA  
WNHLDGPIFAKVECMNKLQDLRGNYFEGQLPVCLGNLKLRLVLDLSNQLSGNL  
ASNSDESLEYLSNNEFFGFSNPLANLTKLVFRJSSSEMLOVETESNWLKPF  
QLTVAALPFCGLKIPNLYQTNRLVLDLSNRLSGDIPWLLNENHNGFQGNLPSSMG  
SFTIPIPTVHKLQVLDFAANDITGVLPDNIHNLRLHNGFQGNLPSSMG  
EMNDISPLDSVYNFSGELPSRLTGCFSITLQLSHNPSGPIPIQTRLTSLVLR  
MNNLTGTEIGVGLRTVLNLSIFDASNNRLTGLISSIPDSSHLLMLLSNNLLEGT  
LPPSLAHLNLDLSNGLSDLPSSVSNMGIKIFLNNSFTGPIPLVTLLENAY  
ILDRNNKSGSIPOVNTGKMITLLRGNLNLSPKRLCDLTSIRLLDLSNKLNG  
VLPCLNHLSTELGELGSGFQSEISFGDSLQMEFYRSTFLVDEFMYYDSYMKIE  
IEFAKQRYDFSGLDLYNVLGDLNLSNLSGVIPEALGDLKRLAINLSRNLSSSI  
PANFKLKDIIESDGLTYNMOGNIHQTLNLTSLAVFNVSFNLSGLIIPQGGFNTFN  
DNYLGNPLLCGTPDRSCGKNTKEADNGEESDDDDDEAADIWVLYWTGTSTY  
AIALIGLVLWMCDFCPRRTWRLCIVDAFIASGSMFS"  
/gene="F9E11.7"  
/notes="similar to GI:1184077 from [Lycopodium  
pimpinellifolium]"  
join(<11609..12002,12342..12497,12658..12732,12853..13008,  
13095..13169,13288..13392,13454..>15699)  
/gene="F9E11.7"  
/join(11609..12002,12342..12497,12658..12732,12853..13008,  
13095..13169,13288..13392,13454..15699)  
/gene="F9E11.7"  
/codon start=1  
/product="disease resistance protein, putative;  
11609-15699"  
/protein\_id="AAG51873.1"  
/db\_xref="GI:12323814"  
/translation="MEGKFLGQYLICVILLGLQHGYSKCIERKALLEKALFLIP  
LNAGEMDNVLSWNTDTSKCCQMMGVCEKGRITNIAFGIGFIENPLNLSLLH  
PFEDVRSLDSSRSRSCDCGFLGDDVEGYSLSLRNLNLEILDSSHRFNNSIFPPL  
NAATSLTLFTLYNNMHSPLVKEFKDLTNLEHLDLGRNFGSIPTQDYNLSRRFRK  
LEILDSDNLFNRSRIFPFLNSATSLKSLGNNMGPPPAKELRDLTNYVELLDLSRN  
RFGSIIPVALFALRKALDLDSDNESSSVELQKFAKTPGSCPCWKNMBEELSL  
NKLACGPFCLSLTGLRLVLDLSNQLTGNVPSALANLESLEVLSPGNFEGFSL  
GLLANLSKLVRLDSQNSLEVEFETSWRPFQVLVIALRSNLEKVPFLHQLKDL  
HVDLSQTHGPNFWLLENNTKLVLLQNNSTFTSQPKSAHNLFLNVSNNFN  
HLFLQNGFWLPHVCLYNGFQNLPSLDNMSIEFLDLSHNRPHGKLPRLFLK  
GCYNLTILKLSHNLKSGVEPEAAFNLMWMDNMLFTGNIGKFRSLPSLNVLDI  
SNKKLTGIVPSWIGERGLFALQLSNNMGEIPTSLEFNISYQLQLDLSNRLSGDIP  
PHVSIYHGAVLLQNNLSGVIPTDILLNVLDRNRLSGNLPEFINTONISILL  
LRGNFTGQIPHQPCSUNLOLLOLNNKNGSIPLCSLNTSFLGRKDDSYRVDVPS  
RFGTAKDPVTFVESLMDIDFNVMNETNSQTKIEPATKRYDAYNGMNLKGLMDLSE  
NELSGTPEVLGGVLEALNLSHNLGVSFGLKSVESLDLSFNRLQGGPIQ  
LTDMSILAVNVSYNLSGVIPOGRQNTFETQYFGNPLLCGKSIDISCAENNFHT  
DNGVEADESTVMSFYWSFVAAYVITLIGLASLSDPSWRAWFIVDAFVLKVRN  
MLWQNTAGTKSVLVLSSPSVVVLLKPPALFKHRT"  
complement(16044..20090)  
/gene="F9E11.1"  
complement(join(<16044..16959,17177..17312,17406..19292,  
19382..19452,19955..>20090))  
/gene="F9E11.1"

gene  
complement(join(16103..16959,17177..17312,17406..19292,  
19382..19452,19955..20090))  
/gene="F9E11.1"  
/codon start=1  
/product="unknown protein; 20090-16103"  
/protein\_id="AAG51874.1"  
/db\_xref="GI:12323815"  
/translation="MAAKLLHSLADSDADLQKQIGCMNGIQIPDRHHVLTGRKSLT  
LGNNAININERDSVDITYQKQETDSNIGNVKERRVSTESSRVSFSSSSSSSS  
SFNGVQPDASAYDRANFQESPTDPENTEGFSGHLGLDLRDVVRDMSREARGLL  
SPTMTREEVRSRREDSPRYGLKQSTPMOLNENFVRLARLRETSHQYNELGMKDA  
PRYSVDHTLKSQKLELPSLDSRERATNNSVDPSKLSLFSFSSSCSSSKK  
RPPSVAKUMGLTLPGLGRDIIHQFLNKTNISQNDPSPRSRLKRNLRNLRIFS  
PSSPSLGDPAKSPARNSDLFVYKQSLNTRFPVEPAPWKAHNRVLRQKQASMPVKAK  
PYEAPNFPPTVYSEMERLNDLEFKHSGDKRLAKQILLESMSQSLFDTEKQOQSTNF  
AVQDYERENSATSNHMSRTRVSSSSNQVQSPVIMPAKLVKAGIPASSLIP  
IHSHTGKIRREKDDKTSASNRVTDCSPGNVRAESCTSSFDKKSQDSNRVRS  
SKPQOVKESASKSGSVSPRLQKLEVDKSRPPTPDSSKSRKPSNQQLVESTS  
PGRRRPKQSLQVDDQSLQSNBSRSTSSHGICTQSTSEASACVEKSTEADGGKSP  
SVTEAKAVVSNLMQNKSPFSEGLSANLSLVALEHSPSPISLDASTRETEPSV  
KTQADHIVGNVAHDFQDCEQMPNPFSETTSFSPENRKLQNVHVLQKLRH  
LNSHDEASQDYIASLCENADPTTHRYISEILLASGLLRLDGLSGLTTPQLHPSGH  
INPELFFVLEQTKGSTTHLLHKEKSKVLEKLNKLVFDLVNLEILVEKLSVEAT  
NLMKSYAKVTKAVSAQQLKELCSAITEITQKQATKRSNFLLSEDDFLKSTLAD  
VTIRSGNADFSGEMSLVDLVERLVKOLVNEIVHAEISRLQAKSGRRRLTADQ"  
complement(23758..26726)  
/gene="F9E11.8"  
/notes="predicted by Genemark.hmm"  
complement(join(<23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..>26726))  
/gene="F9E11.8"  
complement(join(23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..26726))  
/gene="F9E11.8"  
/codon start=1  
/product="hypothetical protein; 26726-23758"

mRNA  
Query Match 87.0%; Score 17.4; DB 15; Length 39601;  
Best Local Similarity 94.7%; Pred. No. 1.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAG 19  
|||||  
Db 1623 CAAGTAAATGCAGACAG 1605

RESULT 50  
AC020579/c  
LOCUS  
DEFINITION  
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,  
complete sequence.  
AC020579  
VERSION  
AC020579.5 GI:12324896  
KEYWORDS  
HTG.  
SOURCE  
Arabidopsis thaliana (thale cress)  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses.  
REFERENCE  
1 (bases 1 to 50821)  
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,  
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Uterback,T.R.,  
Barnstead,M.E., Bowman,C.B., White,O., Nierman,W.C. and Fraser,C.M.  
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence  
Unpublished  
2 (bases 1 to 50821)  
Lin,X. and Kaul,S.  
Direct Submission  
Submitted (05-JAN-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, xlin@igr.org  
3 (bases 1 to 50821)

AUTHORS TITLE JOURNAL COMMENT	Town,C.D. and Kaul,S. Direct Submission Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org On Jan 19, 2001 this sequence version replaced gi:12280787. Address all correspondence to:at@tigr.org	BAC clone F1017 is from Arabidopsis thaliana chromosome 1 clone. The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mt.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/sof/laab/glimmerm.htm/glimmerm.html, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).	gene mRNA CDS	gene mRNA CDS	gene mRNA CDS	repeat_region repeat_region gene mRNA CDS
FEATURES source	Location/Qualifiers 1..50821 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="F1017" /ecotype="Columbia" complement(9837..11757) /gene="F1017.1" /note="similar to GB:AAC49376 from [Glycine max]" complement(join(<9837..10159,10243..10343,10438..10485, 10576..10693,10800..10995,11127..11248,11404..>11757)) /gene="F1017.1" complement(join(10052..10159,10243..10343,10438..10485, 10576..10693,10800..10995,11127..11248,11404..11757)) /gene="F1017.1" /codon_start=1 /product="putative choline kinase; 11757-10052" /protein_id="AAG52400.1" /db_xref="GI:12324898" /translation="MGTEKNYENKQYRLPREVKALQATASEWEDVIDSKALQVPL KGAMTNEVQIKWPTREKPSKVLVRIYGEVEIFFDREDIRTEFPMSKHGQPLL LGRFNGRIEELHARTLSACDLRDEISRIATRMKEPHGEMPAKALLMDRLRN WLTKCKRLASPEAKSFLDVMEMINLEKSLFDNDENIGCHNDLQGNIMMDEET KAITIDYECYNVPAYDIANHFCMAADYHTETPHIMDYSKIPEVERQRLKTYM SYSDKPSDTMYKKLEDEVEKYLASHLWGLWGIISEHVNEIDFDYMEYARQRFQY WLTKRPLLAASEHK" complement(12470..12501) /rpt_family="AT-rich" complement(13045..13067) /rpt_family="(A)n" 13093..16423 /gene="F1017.2" /note="identical to heat shock protein 101 GB:AAF26423 from [Arabidopsis thaliana]" join(<13093..14242,14327..14508,14649..15274,15359..15836, 15941..16423) /gene="F1017.2" join(13093..14242,14327..14508,14649..15274,15359..15836, 15941..16240) /gene="F1017.2" /codon_start=1	gene mRNA CDS	gene mRNA CDS	gene mRNA CDS	repeat_region repeat_region gene mRNA CDS	
	/product="heat shock protein 101; 13093-16240" /protein_id="AAG52410.1" /db_xref="GI:12324908" /translation="MNPEKFTKNTETATTAHELVANAGHAQFTPLHLIAGALISDPTG IFPQAISSAGGENAAQSAERVINQALKLPSPDPDIPASSSLIKVIRRAQAQKS IDRDLAVDQILMGLLEDQIRDLNNEVGATARVKSEVEKRGKGVESASGDTN FOALKTYGRDILVEOAGKLDVIGRDEIRRVIRLSRRTKNPNVLIGEPGQKTAFFE GLAQRIYKGDVPSNLTDVRLISLDMGLVAGAKYRGGBFEERLKSVLKEVEDEAGKVL FIDIEHLVIGAGKTEGSDANLFPKMLARGQLACIGATLLEERKYEKKAADAEERRF QQVYVAFSPDPTISILRGKBEKYEHHGVRIQDRALINAAQLSARYITGRLHPDKAI DLDEACANVRVQLDSQPEEDINLERKMQLEIELHALEREKDKASAKARLLEVRKEL DRDLQPLTMKYRKEKERIDEIRLQKREELMFSLQAEARRYDLAARADRLYGAIQ EVEGAIQLEGTSEENMLTENVPHEIABVSWRTGIPVTRLQONEKERLIGLAIQ LHKRVQONQAVNVAEILRSRAGLGRPOQTSGLFLGPTGVGKTEKALAEOLFL DDENLVRIDMSEYMEQSVSRLLGAPPGYVGEEGQLTEAVRRRPYCVILFUEVEK AHVAVFNTLLQVLDGRLTDGQGRITDFRNSVIITSNLGAELHLAGUTGKVTUEVAR DCVNRVRFKFRPELNLDEIVVDFPLSHDQLRKVARLQMKDVAVRLAERAGALAVT DAADYTLAESYDPVYGARPIRRMEKKVVTLELSMVVREEDENSTVYIDAGAGDLV YRVESGGLVDASTGCKSDVLIHIANGPKRSDAAQAVKMRTEIEDDDNEEMIED" complement(16481..17587) /gene="F1017.3" /note="similar to unknown protein GB:AAD21437 from [Arabidopsis thaliana]" complement(join(<16481..17251,17318..>17587)) /gene="F1017.3" complement(join(16481..17251,17318..17587)) /gene="F1017.3" /codon_start=1 /product="unknown protein; 17587-16481" /protein_id="AAG52411.1" /db_xref="GI:12324909" /translation="MAIGPSSVIIIGIILATYQSLIKPPPNLCSGSPGGPITAPR IKLXDRHLAYKEYLEPREKAKHIVFIHSDSCRHDAVFAITLLSPDLVQBRGYMVS FDKPGYGESDPDPTPKSLADIEELADQLSGSKFYVIQKSMGQAAMGCKLTYPH RLAGVTLVAPVNYWRNLPLNIETEGNLQQRDQWVAHVAPMLIYWNNTQNF PGSSVNRDGVLSQPKDITILKLGSSRKPLAEVRQGIHESINRDMIVGFGNWEFD PLELENPLNREGSVHLWQGDDEMLVPVTLQRYTADKLPLWHLHYEVAGCGHFFPLAKG VVDIIVKTLNH" complement(18687..21119) /gene="F1017.4" /note="similar to unknown protein GB:AAD21437 from [Arabidopsis thaliana]" complement(join(<18687..19284,20351..20547,20799..>21119)) /gene="F1017.4" complement(join(18687..19284,20351..20547,20799..21119)) /gene="F1017.4" /codon_start=1 /product="unknown protein; 21119-18687" /protein_id="AAG52412.1" /db_xref="GI:12324910" /translation="MASDSRSKSLKPSRSKFLPSPVIVIVIGIIVALTQSKLKPP QPKLGGSSGPPITAPRIKLDQGRYLAYEHGLPREKANRKIVFIHSDCCRHDAVFA TLLSPDLVEELGVYMSFDRPGYCESDPHPSRTPSLSVSDIEELDQLSGSKFYVIG KSMGQAAMGCKLNYIPIHRLAGVTLVAPVNYWRNLPLNVESTGFNFQQRDQWV RVAVAPMLIYWNNTQKWPFGSSIANRDSLSQSDROIISKRGYTRKPHMAEVRQGI HESINRDMIVGFGNWEFDPLDNDPFLNNEGFLWQGDDEMLVPVTLQRYLAHLPLW VHYHVPKSGHFFHFTKGVNDNIVTLLTTDTDIRQ" complement(21829..23197) /gene="F1017.5" /note="similar to unknown protein GB:AAD21437 from [Arabidopsis thaliana]" complement(join(<21829..22626,22877..>23197)) /gene="F1017.5" complement(join(21829..22626,22877..23197)) /gene="F1017.5" /codon_start=1 /product="unknown protein; 23197-21829" /protein_id="AAG52399.1" /db_xref="GI:12324897" /translation="MTRDSRSNLSKMPWRKFLPSPVIVIVIGIIVALTQSKLKPP PPLKGGSSGPPITAPRIKLDQGRYLAYEHGLPREKANRKIVFIHSDCCRHDAVFA TLLSPDLVEELGVYMSFDRPGYCESDPHPSRTPSLSVSDIEELDQLSGSKFYVIG YSMGQAAMGCKLNYIPIHRLAGVTLVAPVNYWRNLPLNVESTGFNFQQRDQWV	gene mRNA CDS	gene mRNA CDS	gene mRNA CDS		

```

AHYTPWLIYWNWTKQWFGSSIANRHSLLAQFDKIISKLGSSRKPWHAEVROQGIH
ESINRDMIVGFNGWEPGLDLENPLNKGSGVHLWQGDMDLVPKLORYLAHQLPWW
HYHEVPRSGHFHYHTKGVDVDDIVKSLTSDVRSRVCII
23489..24554
/genes="F1017.6"
/notes="similar to ribosomal protein L33B GB:NP_014877 from
[Saccharomyces cerevisiae]"
join(<23489..23520,23750..23785,24176..24404,24499..24554)
/genes="F1017.6"
join(23489..23520,23750..23785,24176..24404,24499..24540)
/genes="F1017.6"
/codon_start=1
/product="putative ribosomal protein; 23489-24540"
/protein_id="AAG52401.1"
/db_xref="GI:12324899"
/translation="MVKGQGERVLYVGTILGVYKRSKSNYPNTSLVQIRGVNTOE
ENVYKGRMAYIIKAKTKNGSHYKCIWKTFRPHNGSGVVRKFTSNLPFKSMGSR
VRVFMVPSNI
25509..29950
/genes="F1017.7"
/notes="similar to GB:K1AA0361 from [Homo sapiens]"
join(<25509..25911,25995..26563,26663..29047,
29144..>29950)
/genes="F1017.7"
/genes="F1017.7"
/codon_start=1
/product="putative phosphoribosylformylglycinamide
synthase; 25509-29950"
/protein_id="AAG52403.1"
/db_xref="GI:12324901"

Query Match      87.0%; Score 17.4; DB 15; Length 50821;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
Db 47324 CAAGTAAATGCAGACAG 47306

RESULT 51
LOCUS AC019512 51405 bp DNA linear HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC019512
VERSION AC019512.1 GI:6665385
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 51405)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10210037 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1..51405
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match      87.0%; Score 17.4; DB 14; Length 51405;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;

```

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AAGTAAATGCAGAAACAG 20
Db 18370 AGGTAATGCAGAAACAG 18388

RESULT 52
LOCUS AC100240/c 54288 bp DNA linear HTG 22-NOV-2001
DEFINITION Mus musculus clone RP23-64E10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100240
VERSION AC100240.1 GI:17047606
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 54288)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-64E10
Unpublished
2 (bases 1 to 54288)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chospel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McSwan, P., McKernan, K., McPheeters, R., Meldrin, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Olivier, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zairoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14437
Center clone name: 64_E_10
-----
* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 694: contig of 694 bp in length

```

```

* 695 794: gap of 100 bp
* 795 1507: contig of 713 bp in length
* 1508 1507: gap of 100 bp
* 1608 2279: contig of 672 bp in length
* 2280 2379: gap of 100 bp
* 3088 3087: contig of 708 bp in length
* 3188 3187: gap of 100 bp
* 3898 3897: contig of 710 bp in length
* 3998 4705: contig of 708 bp in length
* 4806 4805: gap of 100 bp
* 5591 5590: gap of 100 bp
* 6283 6282: contig of 692 bp in length
* 6383 7088: contig of 706 bp in length
* 7089 7188: gap of 100 bp
* 7872 7871: contig of 683 bp in length
* 7972 7971: gap of 100 bp
* 8652 8651: contig of 680 bp in length
* 8752 8751: gap of 100 bp
* 9461 9460: contig of 709 bp in length
* 9561 10265: contig of 705 bp in length
* 10366 11085: contig of 720 bp in length
* 11086 11887: gap of 100 bp
* 11887 11887: contig of 702 bp in length
* 11888 11987: gap of 100 bp
* 11988 12688: contig of 701 bp in length
* 12689 12788: gap of 100 bp
* 12789 13485: contig of 697 bp in length
* 13486 13585: gap of 100 bp
* 13586 14268: contig of 683 bp in length
* 14269 14368: gap of 100 bp
* 14369 15078: contig of 710 bp in length
* 15079 15178: gap of 100 bp
* 15179 15890: contig of 712 bp in length
* 15891 15990: gap of 100 bp
* 15991 16692: contig of 702 bp in length
* 16693 16792: gap of 100 bp
* 16793 17465: contig of 673 bp in length
* 17466 17565: gap of 100 bp
* 17566 18237: contig of 672 bp in length
* 18238 18337: gap of 100 bp
* 18338 19029: contig of 692 bp in length
* 19030 19129: gap of 100 bp
* 19130 19828: contig of 699 bp in length
* 19829 19928: gap of 100 bp
* 19929 20610: contig of 682 bp in length
* 20611 20710: gap of 100 bp
* 20711 21415: contig of 705 bp in length
* 21416 21515: gap of 100 bp
* 21516 22224: contig of 709 bp in length
* 22225 22324: gap of 100 bp
* 22325 23024: contig of 700 bp in length
* 23025 23124: gap of 100 bp
* 23125 23825: contig of 701 bp in length
* 23826 23925: gap of 100 bp
* 23926 24628: contig of 703 bp in length
* 24629 24728: gap of 100 bp
* 24729 25444: contig of 716 bp in length
* 25445 25544: gap of 100 bp
* 25545 26274: contig of 730 bp in length
* 26275 26374: gap of 100 bp
* 26375 27066: contig of 692 bp in length
* 27067 27166: gap of 100 bp
* 27167 27852: contig of 686 bp in length
* 27853 27952: gap of 100 bp
* 27953 28662: contig of 710 bp in length
* 28663 28762: gap of 100 bp
* 28763 29463: contig of 701 bp in length
* 29464 29563: gap of 100 bp

```

```

* 29564 30266: contig of 703 bp in length
* 30267 30366: gap of 100 bp
* 30367 31070: contig of 704 bp in length
* 31071 31170: gap of 100 bp
* 31171 31885: contig of 715 bp in length
* 31886 31885: gap of 100 bp
* 31886 32669: contig of 684 bp in length
* 32670 32769: gap of 100 bp
* 32770 33451: contig of 682 bp in length
* 33452 33551: gap of 100 bp
* 33552 34243: contig of 692 bp in length
* 34244 34343: gap of 100 bp
* 34344 35047: contig of 704 bp in length
* 35048 35147: gap of 100 bp
* 35148 35858: contig of 711 bp in length
* 35859 35958: gap of 100 bp
* 35959 36685: contig of 727 bp in length
* 36686 36785: gap of 100 bp
* 36786 37495: contig of 710 bp in length
* 37496 37595: gap of 100 bp
* 37596 38294: contig of 699 bp in length
* 38295 38394: gap of 100 bp
* 38395 39096: contig of 702 bp in length
* 39097 39196: gap of 100 bp
* 39197 39882: contig of 686 bp in length
* 39883 39882: gap of 100 bp
* 39883 40665: contig of 683 bp in length
* 40666 40765: gap of 100 bp
* 40766 41474: contig of 709 bp in length
* 41475 41574: gap of 100 bp
* 41575 42253: contig of 679 bp in length
* 42254 42353: gap of 100 bp
* 42354 43048: contig of 695 bp in length
* 43049 43148: gap of 100 bp
* 43149 43862: contig of 714 bp in length
* 43863 43962: gap of 100 bp
* 43963 44643: contig of 681 bp in length
* 44644 44743: gap of 100 bp
* 44744 45461: contig of 718 bp in length
* 45462 45561: gap of 100 bp
* 45562 46252: contig of 691 bp in length
* 46253 46353: gap of 100 bp
* 46353 47059: contig of 707 bp in length
* 47060 47159: gap of 100 bp
* 47160 47864: contig of 705 bp in length
* 47865 47964: gap of 100 bp
* 47965 48676: contig of 712 bp in length
* 48677 48776: gap of 100 bp
* 48777 49479: contig of 703 bp in length
* 49480 49579: gap of 100 bp
* 49580 50279: contig of 700 bp in length
* 50280 50379: gap of 100 bp
* 50380 51068: contig of 689 bp in length
* 51069 51168: gap of 100 bp
* 51169 51879: contig of 711 bp in length
* 51880 51979: gap of 100 bp
* 51980 52689: contig of 710 bp in length
* 52690 52789: gap of 100 bp
* 52790 53480: contig of 691 bp in length
* 53481 53580: gap of 100 bp
* 53581 54288: contig of 708 bp in length.

```

FEATURES source  
Location/Qualifiers 1. 54288

Query Match 87.0%; Score 17.4; DB 14; Length 54288;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAACAGG 20  
DB 21776 AAGTAAATGCAGAACAGG 21758

RESULT 53  
AC101468/c  
LOCUS AC101468 63540 bp DNA linear HTG 23-NOV-2001  
DEFINITION Mus musculus clone RP23-187111, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC101468  
VERSION AC101468.1 GI:17060243  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
1 (bases 1 to 63540)  
Biren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP23-187111  
Unpublished  
2 (bases 1 to 63540)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L16812  
Center clone name: 187\_I\_11  
-----  
\* NOTE: This record contains 80 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 755: contig of 755 bp in length  
\* 756 855: gap of 100 bp  
\* 856 1570: contig of 715 bp in length  
\* 1571 1670: gap of 100 bp  
\* 1671 2371: contig of 701 bp in length  
\* 2372 2471: gap of 100 bp  
\* 2472 3196: contig of 725 bp in length  
\* 3197 3296: gap of 100 bp  
\* 3297 3982: contig of 686 bp in length  
\* 3983 4082: gap of 100 bp  
\* 4083 4765: contig of 683 bp in length  
\* 4766 5537: gap of 100 bp  
\* 5538 5637: contig of 672 bp in length  
\* 5638 6308: contig of 671 bp in length  
\* 6309 7094: contig of 686 bp in length  
\* 7095 7194: gap of 100 bp  
\* 7195 7883: contig of 689 bp in length  
\* 7884 8659: contig of 676 bp in length  
\* 8660 9468: gap of 100 bp  
\* 9469 10282: contig of 715 bp in length  
\* 10283 10382: gap of 100 bp  
\* 10383 11099: contig of 717 bp in length  
\* 11100 11903: contig of 704 bp in length  
\* 11904 12696: contig of 693 bp in length  
\* 12697 13468: contig of 672 bp in length  
\* 13469 13568: gap of 100 bp  
\* 13569 14233: contig of 665 bp in length  
\* 14234 14333: gap of 100 bp  
\* 14334 15003: contig of 670 bp in length  
\* 15004 15801: contig of 698 bp in length  
\* 15802 15901: gap of 100 bp  
\* 15902 16605: contig of 704 bp in length  
\* 16606 16705: gap of 100 bp  
\* 16706 17423: contig of 718 bp in length  
\* 17424 18291: gap of 100 bp  
\* 18292 18391: gap of 100 bp  
\* 18392 19109: contig of 718 bp in length  
\* 19110 19209: gap of 100 bp  
\* 19210 19224: contig of 715 bp in length  
\* 19225 20024: gap of 100 bp  
\* 20025 20739: contig of 715 bp in length  
\* 20740 20839: gap of 100 bp  
\* 20840 21524: contig of 685 bp in length  
\* 21525 21624: gap of 100 bp  
\* 21625 22304: contig of 680 bp in length  
\* 22305 22404: gap of 100 bp  
\* 22405 23089: contig of 685 bp in length  
\* 23090 23189: gap of 100 bp  
\* 23190 23881: contig of 692 bp in length  
\* 23882 23981: gap of 100 bp  
\* 23982 24698: contig of 717 bp in length  
\* 24699 24798: gap of 100 bp  
\* 24799 25426: contig of 628 bp in length  
\* 25427 25526: gap of 100 bp  
\* 25527 26246: contig of 720 bp in length  
\* 26247 26346: gap of 100 bp  
\* 26347 27060: contig of 714 bp in length  
\* 27061 27160: gap of 100 bp  
\* 27161 27866: contig of 706 bp in length  
\* 27867 27966: gap of 100 bp  
\* 27967 28657: contig of 691 bp in length  
\* 28658 28757: gap of 100 bp  
\* 28758 29427: contig of 670 bp in length  
\* 29428 29527: gap of 100 bp  
\* 29528 30214: contig of 687 bp in length  
\* 30215 30314: gap of 100 bp  
\* 30315 31011: contig of 697 bp in length  
\* 31012 31111: gap of 100 bp  
\* 31112 31803: contig of 692 bp in length  
\* 31804 31903: gap of 100 bp  
\* 31904 32609: contig of 706 bp in length

KEYWORDS  
SOURCE  
ORGANISMHTG.  
Homo sapiens (human)  
Homo sapiensEukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.REFERENCE  
1 (bases 1 to 65756)

Tracey, A.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Dec 5, 2001 this sequence version replaced gi:16944115.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

RP11-157L10 is from the library RP11-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VSECTOR: pBACE3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality &gt;=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one subclone; and the assembly was confirmed by restriction digest,

except on the rare occasion of the clone being a YAC.

Location/Qualifiers

1. .65756

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP11-157L10"

/clone\_lib="RP11-11.1"

2000

/note="Clone right end: RP11-57022"

complement(6159..6449)

/locus\_tag="RP11-157L10.1-001"

/pseudo

complement(6159..6449)

/locus\_tag="RP11-157L10.1-001"

/note="match: proteins: AAL87001 P49207 Q29223 Q9BUZ2

Q9D1R9"

/pseudo

/codon\_start=1

10497..10665

/note="Single clone region. Sequence from reads from a

short insert library derived from a single pUC clone.

Restriction digest data confirm the assembly."

48524

/note="Clone\_left\_end: RP11-62E18"

63757

/note="Clone\_left\_end: RP1-257A15"

-----

ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 65756;

Best Local Similarity 94.7%; Pred. No. 1.2e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAGG 20

\* 32610 32709: gap of 100 bp  
\* 33419: contig of 710 bp in length  
\* 33420 33519: gap of 100 bp  
\* 33520 34208: contig of 689 bp in length  
\* 34209 34308: gap of 100 bp  
\* 34309 34954: contig of 646 bp in length  
\* 34955 35054: gap of 100 bp  
\* 35055 35734: contig of 680 bp in length  
\* 35735 35834: gap of 100 bp  
\* 35835 36507: contig of 673 bp in length  
\* 36508 36607: gap of 100 bp  
\* 36608 37273: contig of 666 bp in length  
\* 37274 37373: gap of 100 bp  
\* 37374 38052: contig of 679 bp in length  
\* 38053 38152: gap of 100 bp  
\* 38153 38859: contig of 707 bp in length  
\* 38860 38959: gap of 100 bp  
\* 38960 39664: contig of 705 bp in length  
\* 39665 39764: gap of 100 bp  
\* 39765 40448: contig of 684 bp in length  
\* 40449 40548: gap of 100 bp  
\* 40549 41272: contig of 724 bp in length  
\* 41273 41372: gap of 100 bp  
\* 41373 42086: contig of 714 bp in length  
\* 42087 42186: gap of 100 bp  
\* 42187 42866: contig of 680 bp in length  
\* 42867 42966: gap of 100 bp  
\* 42967 43660: contig of 694 bp in length  
\* 43661 43760: gap of 100 bp  
\* 43761 44449: contig of 689 bp in length  
\* 44450 44549: gap of 100 bp  
\* 44550 45222: contig of 673 bp in length  
\* 45223 45322: gap of 100 bp  
\* 45323 46002: contig of 680 bp in length  
\* 46003 46102: gap of 100 bp  
\* 46103 46788: contig of 686 bp in length  
\* 46789 46888: gap of 100 bp  
\* 46889 47606: contig of 718 bp in length  
\* 47607 47706: gap of 100 bp  
\* 47707 48430: contig of 724 bp in length  
\* 48431 48530: gap of 100 bp  
\* 48531 49241: contig of 711 bp in length  
\* 49242 49341: gap of 100 bp  
\* 49342 50060: contig of 719 bp in length  
\* 50061 50160: gap of 100 bp  
\* 50161 50868: contig of 708 bp in length  
\* 50869 50969: gap of 100 bp  
\* 50969 51649: contig of 681 bp in length  
\* 51650 51750: gap of 100 bp  
\* 51750 52427: contig of 678 bp in length  
\* 52428 52527: gap of 100 bp  
\* 52528 53202: contig of 675 bp in length  
\* 53203 53302: gap of 100 bp  
\* 53303 53957: contig of 655 bp in length  
\* 53958 54057: gap of 100 bp  
\* 54058 54747: contig of 690 bp in length

Query Match 87.0%; Score 17.4; DB 14; Length 63540;

Best Local Similarity 94.7%; Pred. No. 1.2e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19

|||||

Db 9262 CAAGTAAATTCAGAAACAG 9244

## RESULT 54

AL590286/c

LOCUS

AL590286 65756 bp DNA linear PRI 18-MAY-2005

DEFINITION Human DNA sequence from clone RP11-157L10 on chromosome 6 Contains

a ribosomal protein L34 (RPL34) pseudogene, complete sequence.

ACCESSION AL590286

VERSION AL590286.12 GI:17384111

Db	537		537	AAGTAAATACAGAACAGG	519	
RESULT 55						
AC003096/c						
LOCUS						
DEFINITION						
AC003096						
AC003096						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
AC003096						
69817 bp	DNA	linear	PLN 27-FEB-2002			
Arabidopsis thaliana chromosome 2 clone T29F13 map ve016, complete						
sequence.						
AC003096						
HTG.						
AC003096.3	GI:20197018					
Arabidopsis thaliana (thale cress)						
Arabidopsis thaliana						
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;						
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.						
1 (bases 1 to 69817)						
Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,						
Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,						
Somerville,C.R. and Venter,J.C.						
Unpublished						
2 (bases 1 to 69817)						
Lin.X.						
Direct Submission						
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712						
Medical Center Dr., Rockville, MD 20850, USA						
3 (bases 1 to 69817)						
Town,C.D. and Kaul,S.						
Direct Submission						
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712						
Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org						
On Apr 18, 2002 this sequence version replaced gi:6598389.						
COMMENT						
FEATURES						
source						
1. 69817						
/organism="Arabidopsis thaliana"						
/mol_type="genomic DNA"						
/db_xref="taxon:3702"						
/chromosome="2"						
/map="ve016"						
/clone="T29F13"						
/ecotype="Columbia"						
complement(1..1207)						
/notes="overlap with BAC clone F1913 (AC004238:1..1207)"						
complement(4397..6665)						
/gene="At2g34770"						
/notes="synonym: T29F13.2; identical to GB:AF021804;						
supported by cDNA: gi_15215595.gb AY050326.1"						
complement(join(<4397..4589,4692..4838,4916..5021,						
5155..5351,5460..5552,5705..5770,6641..6665))						
/gene="At2g34770"						
complement(join(4470..4589,4692..4838,4916..5021,						
5155..5351,5460..5552,5705..5755))						
/gene="At2g34770"						
/codon_start=1						
/product="fatty acid hydroxylase (FAH1)"						
/protein_id="AAC16270.1"						
/db_xref="GI:3132481"						
/translation="MWAQGFVDLKKPLVFQVGHGEDVEVWHQFIATKEGRRFFQS						
DFWEFLTLVMAVPVILWLVVWCISRSVSMGCSLPEIPIVMVGIFWTFFEYVLH						
RFVFIHTKSYNGNTAHYLIHGCHKHMDHLVFPPTAILCPFWNIKAISTP						
STAPALGGGMLGVMDVTHYLLHQAQPTPTVTKNLKXVHLNHHFRIQDKGFGITSS						
LWDVFGTLPTITKAPKEQ"						
7682..7862						
/gene="At2g34760"						
/note="synonym: T29F13.3; pseudogene, tubulin beta chain"						
/pseudo						
8739..12082						
/gene="At2g34750"						
/note="synonym: T29F13.4"						
join(<8739..8867,8989..9078,9171..9242,9347..9418,						
9505..9741,9833..9979,10031..10135,10253..10378,						
10487..10555,10667..10788,11071..11158,11340..11453,						
11547..11731,11830..12082)						
/gene="At2g34750"						
/codon_start=1						
/product="unknown protein"						
/protein_id="AAC16259.1"						
/db_xref="GI:3132470"						
/translation="MGAVELMSDPSSLCTVENYVDNVDLSDTQLVQTVTKALTSVKTG						
DSDLXSEWVGWARDIKFEPDQDVVAQLETVLKALSGAVACIDVLHGHOKLLGALPMK						
LWDHPVDNALNLVLSIATSGTKLDSCLNMLVSNFVPPVQVNNVSHSLNKKKI						
DVLSRVHAALLKISILVPLTSRPLVQMPHQPMKHKKDHSIIVIVESLKLUESSIG						
GTMGSMILGWYMERLDVSRQNNLIQLRSLIEIEMDDIPQDSSRGFMFMBELDAE						
QVNDGDLVPGLQKDTSDGSIVSKLLDKLMVAPEHLESQNDGRLDQVFESLRSKI						
FENFTLNTYKSKFTQFLIFYACSLDPENCVKVFASKLVEIPLSSNKHVATROASLRKI						
DEVCYCTCTDTRPEAHQIFPSCQALIMVYLCFMRSLIDVPRFSQLTPLSILM						
HKLNPLMVLPSVVAEFURQAKGEGFIVSDSFIQDLSLSLSRAFGFERLDTFFP						
FDPCILKSSNSFISPNFIYGNVWKATYDEDDDDNDAEVIIVNGDESDSDEADLDTAL						
NKMSITPKHSFKNMERDLRLMPSRIIRPSTSPESL"						
complement(11877..11962)						
/rpt_family="CAT'n"						
complement(12226..13210)						
/gene="At2g34740"						
/note="synonym: T29F13.5"						
complement(join(<12226..12375,12458..12612,12713..12989,						
13073..13210))						
/gene="At2g34740"						
/codon_start=1						
/product="putative protein phosphatase 2C"						
/protein_id="AAC16260.1"						
/db_xref="GI:3132471"						
/translation="MEDFIVADTKTVKGNHNLGLYAIIPDGHSGSDVADYLVQNLFDNII						
SQPDFRNPKAIKRAYKSTDDYILQNVVPRGGSTAVTAIVIDGKIVVAVNGDSRA						
ILCRSDVVVKQITVDHPDKERDLVSKGKGVSKPGNVPVVDGLAMTRAFDGLGK						
EHISVTPNIEIAEIHDDTKFLILASDGLWKVMSNDEVDQIKKRGNAEAAKMLIDKA						
LARGSKDDISCVVVSFLQWID"						
13276..13359						
/rpt_family="CAT'n"						
complement(14252..14307)						
/rpt_family="AT rich"						
complement(15165..17920)						
/gene="At2g34730"						
/note="synonym: T29F13.6"						
complement(join(<15165..15344,15503..15629,15738..>17920))						
/gene="At2g34730"						
complement(join(15165..15344,15503..15629,15738..17920))						
/gene="At2g34730"						
/codon_start=1						
/product="putative myosin heavy chain"						
/protein_id="AAC16261.1"						
/db_xref="GI:3132472"						
/translation="MEEVVVKGSGLEFHDHDTLSSSLQVNGVLKENENPDVDFLEDL						
DSYWDINRLRTISIRGVMTAIESDAABKIAOKDLSELSKIRETLVLLVHVS						
ENESSEPHLIDELTQVSSSLKKARQQLMLVLELTNLREYIHNGSGATVDDSL						
GLDSSPHETRTKTVDKMLDSLKILETVLKRKNDMLPSWQEHDFQKEIESAVTS						
VLRSLKDEYEQRLDKQAEFGGNSRLINKEITGLRQLEAEIRKSLFDHENGDEAG						
EVGDRKRVQLHRKMSGLNSVSSWENGHBESSTGLIPEHNETHRMSPDEMINHP						
KIEMNKKRDHDYKIOELTEQCTPKRKYLNLTTERGSFSGVKDKLGAALKKIPFI						
SKLDKLMDEKFPVSGKNDAGLKRLDLSLLNRLKDSLDAAEKMSLSQAEADH						
QELIRKLTDEDSRNEASIEDVYCGFTVFQCIKCTQETDLEHSMLEAYELL						
EDLARKEARKSKEDFEDSCVKSVMMECCSVIYKEAKHKKIYELNLTNLTVEKGTIL						
RSEMDKERLKEBIHRLGLVKEKNLVQENNLATERKKIIEVVSQIQLSQSVER						
QETEIQDKTEALSVSARLEKVKYETKISSIRLELELAKSKMOKRSKQSDLS						
SETKAKETLKKQLSLDVLVPPQLIKGPDILRGLTAEKTKTNSRLKMSQSLSDLS						
HQINEVKGKASTYKQLEKCCDLKKAERVDLGLGDEVTLLDLLEKIYIALDHYSPI						
LKHYPGVSTISVYTFIFISILRLIMRPY"						
19787..19824						
repeat_region						

```
repeat_region /rpt_family="AT rich"
gene complement(20248..20271)
      /rpt_family="(A)n"
      21380..22513
      /gene="At2g34720"
      /note="synonym: T29F13.7"
mRNA join(<21380..21511,21666..21761,21897..21980,22069..22233,
      22310..>22513)
      /gene="At2g34720"
CDS join(21380..21511,21666..21761,21897..21980,22069..22233,
      22310..22513)
      /gene="At2g34720"
      /codon_start=1
      /product="putative CCAAT-binding transcription factor
      subunit"
      /protein_id="AAC16262.1"
      /db_xref="GI:3132473"
      /translation="MFCWSDNNESHAKERPDSOTPOVPSGRSSSEIDTWSVSEPM
      PCFDAYNAKTLWELIIOAKSRDRLSAVLDPAGLGYPPDPYRSVFAQAQYLPHPY
      PVQIQLMGMOQPGFLOCDVVEPVPVNAKYHGLIRRSRAKLEARNRAIKAKKP
      YVHESRHLHAIHRRPGCGRFLNAKKGDKHKEBEATSDENTSASSLSLRSEKLAA
      TSGPNGRS"
      2437..24462
repeat_region /rpt_family="AT rich"
      complement(26798..26850)
repeat_region /rpt_family="AT rich"
      complement(27849..27880)
      /rpt_family="(A)n"
misc_feature 27932..33247
      /note="molecular marker Athb_14"
gene 28199..33208
      /gene="At2g34710"
      /note="synonym: T29F13.8; identical to GB:Y111122"
mRNA join(28199..28770,29333..29492,29584..29699,29785..29875,
      29992..30087,30169..30366,30441..30596,30681..30755,
      30844..30950,31036..31225,31317..31505,31597..31677,
      31760..31900,31992..32305,32404..32518,32604..32753,
      32837..33208)
      /gene="At2g34710"
      /complement(28213..28274)
repeat_region /rpt_family="(GGAG)n"
      join(28538..28770,29333..29492,29584..29699,29785..29875,
      29992..30087,30169..30366,30441..30596,30681..30755,
      30844..30950,31036..31225,31317..31505,31597..31677,
      31760..31900,31992..32305,32404..32518,32604..32753,
      32837..32983)
      /gene="At2g34710"
      /codon_start=1
      /product="homeodomain transcription factor (ATHB-14)"
      /protein_id="AAC16263.1"
      /db_xref="GI:3132474"
      /translation="MMVHMSRDMNMRSPDKLDSGKYVRYTPEQVEALERVYTEC
      PKPSSLRRLQRCPLISNPKOIKVFNORCRKORKEARLQTVNRKLNAMNK
      LLMEENDRLQKOVNLVYENGHKQLHTASCTTTDNCSESVVSGQQOQNPQKH
      QORDANNPGLUSIAEAAELFSLKATGAVDWQVQIGMKPGPDSIGIVAIRNCSGI
      AARACGLVLEPMKVAELIKDRSRLDSVDTLSVPAGNGTIELIYTOYAPT
      LAAARDFTLRYSTCLDGSYVVCERSLTSGTGPSPSNFVRAEMKPSGFLIRPC
      DGGGSLTHIVDVLDAVSPVMPPLVYESSKILAKMTVAALRHVROIQAIQTSGEVQ
      YGGGQPAVLRTFSQRLCGFNDVNGFVDDGWSPMGSDGAEVTVMLNLSPGKFGGS
      QYGNSEFLPSFGSGVLCAKASMLLQNVPPAVLVRFLREHRSWADYGVDAAYAAASLRAS
```

```
DEFINITION Human DNA sequence from clone RP11-309C8 on chromosome X, complete
sequence.
ACCESSION AL662897
VERSION AL662897.6 GI:18307360
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homnidae; Homo.
REFERENCE 1 (bases 1 to 84987)
AUTHORS Howden, P.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT Clone requests: clonerequest@sanger.ac.uk
      On Jan 23, 2002 this sequence version replaced gi:18250572.
      The following abbreviations are used to associate primary accession
      numbers given in the feature table with their source databases:
      Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
      on the WORMPEP database can be found at
      http://www.sanger.ac.uk/Projects/c_elegans/wormpep This sequence
      was generated from part of bacterial clone contigs of human
      chromosome X, constructed by the Sanger Centre Chromosome X Mapping
      Group. Further information can be found at
      http://www.sanger.ac.uk/HGP/ChrX
      RP11-309C8 is from the library RPCI-11.2 constructed by the group
      of Pieter de Jong. For further details see
      http://www.chori.org/bacpac/home.htm
      VECTOR: pBACe3.6
      ----- Genome Center
      Center: Wellcome Trust Sanger Institute
      Center code: SC
      Web site: http://www.sanger.ac.uk
      Contact: vegas@sanger.ac.uk
      -----
      This sequence was finished as follows unless otherwise noted: all
      regions were either double-stranded or sequenced with an alternate
      chemistry or covered by high quality data (i.e., phred quality >=
      30); an attempt was made to resolve all sequencing problems, such
      as compressions and repeats; all regions were covered by at least
      one subclone; and the assembly was confirmed by restriction digest,
      except on the rare occasion of the clone being a YAC.
FEATURES
      source
      1..84987
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="RP11-309C8"
      /clone_lib="RPCI-11.2"
ORIGIN
      Query Match 87.0%; Score 17.4; DB 8; Length 84987;
      Best Local Similarity 94.7%; Pred. No. 1.2e+03;
      Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAAGTAAATCGAGAAACAG 19
      |||||
      64774 CAAGTTAATCGAGAAACAG 64756
Db
      Query Match 87.0%; Score 17.4; DB 15; Length 69817;
      Best Local Similarity 94.7%; Pred. No. 1.2e+03;
      Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
RESULT 57
AC106751
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-164E14, WORKING DRAFT
      AC106751 101526 bp DNA linear HTG 12-JAN-2002
      SEQUENCE, 3 unordered pieces.
ACCESSION AC106751
VERSION AC106751.1 GI:18139301
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```





```
complement (41919..42054), complement (40888..40984),
complement (39325..39414), complement (32086..32224),
complement (30138..30237), complement (28461..28633),
complement (18933..19030), complement (15068..15229),
complement (11745..11959), complement (11074..11335),
complement (10794..10982), complement (9234..9424),
complement (8975..9046), complement (7331..7451),
complement (2360..2470),
complement (AL355472.19:90370..90519),
complement (AL355472.19:82731..82817),
complement (AL355472.19:78189..78385),
complement (AL355472.19:73467..73548),
complement (AL355472.19:70668..70839),
complement (AL355472.19:65932..66123),
complement (AL355472.19:65648..65772),
complement (AL355472.19:64702..64838),
complement (AL355472.19:63599..64031),
/locus_tag="RP5-827C21.5-001",
/product="TAR (HIV) RNA binding protein 1",
/notes="match: ESTs: Em:AA714641.1 Em:AI079700.1
Em:BM989347.1
match: cDNAs: Em:U38847.1"
Join(complement (59831..60761), complement (54372..54469),
complement (52846..52915), complement (49160..49308),
complement (47314..47366), complement (45495..45592),
complement (41919..42054), complement (40888..40984),
complement (39325..39414), complement (32086..32224),
complement (30138..30237), complement (28461..28633),
complement (18933..19030), complement (15068..15229),
complement (11745..11959), complement (11074..11335),
complement (10794..10982), complement (9234..9424),
complement (8975..9046), complement (7331..7451),
complement (2360..2470),
complement (AL355472.19:90370..90519),
complement (AL355472.19:82731..82817),
complement (AL355472.19:78189..78385),
complement (AL355472.19:73467..73548),
complement (AL355472.19:70668..70839),
complement (AL355472.19:65932..66123),
complement (AL355472.19:65648..65772),
complement (AL355472.19:64702..64838),
complement (AL355472.19:63863..64031),
/locus_tag="RP5-827C21.5-001",
/standard_name="ORTHOP00000036048"
/notes="match: proteins: Tr:Q13395"
/codon_start=1
/product="TAR (HIV) RNA binding protein 1"
/protein_id="CAI22828.1"
/db_xref="GI:56204383"
/db_xref="Genew:11568"
/db_xref="GOA:Q13395"
/db_xref="InterPro:IPR001537"
/db_xref="UniProt/TrEMBL:Q13395"
/translation="MEWLAELLSQSDPRALLGALCGEASAEVETLFLQLRLE
DEARGSGAGALPAANEVAAGYVPLRLSLRGPAGDPDSLPQRRHRLVRAAGA
ALRSCVRLAGRLPQLAAALAEALRDLLAGWAPGAEEVLAAGVPCIRPREDGPLL
ERVATVLAALGAGDGDDEAGPAEDAAALVAGRLPVLCQGAALRAVMGGLAAGP
ASLGSGRVEELKLLVLSALAEKLLPEFGDRAGAREAGPDRRCWRFWRTQAGLQGA
DALTKRYALLQRAVEVSAELGADCTCGPQSGNGPSLFWMSERKKDELLKFWENYIL
TMELEGNOIHWIKVLPKLNLFYAVSEENGWCLPHPSWHMCILYKMFSENNKILS
KGVYTHFLYELTKILPSPSEFIIGPLMDALSESLYSRSPQPIGSCSPGLKGL
KQFVYISLPEEIKSFLKFIKRMTRSHWCAPFLLSKALANVRHKAIGIDGL
LALRDVHCTMTHTOILRGAACQYLLOTANMLLDEKVSLSDSVTFLMSLRQESGL
RGTSLWELCMTLWRNYESYFKPSPTCSSTGLHKTSLNAYKVSIVQEVYKSAWETGEN
CMPDPWEAKVLSMLVLLADVDEGMKTOYSGKQRTENVLRIFLPLDLLVMKFSFNAY
MPLLTDLRCLQLLLKLLNLCRLKSSAQDDVESTVLQPFMSTTESISEFLLRLTMN
ELNSVDDLRCHLMLTELNLHNLKWRKGNFIWRVLSLLKNASIQHQLQSDSGQ
EPTQSGYQORVVSMAALAWCEAIQDKPELQDLSHAGLESFLSLQNLQTLQKPHA
EFQSSYAHPLCESSSVSSQGGKIVAYIQHWVWCLSLFKKXHTLITPTGSEI
LEPFLPVMQMPRTLQSALEALTVLSSDQVPLVFLVPLKLTSSSLCIESFDM
AWKLTSSLSNTOLIETWANKAFVQVFDQNKVLTIAAKTKGOAYFKIKIEMYKIMSA
IKTGVPNTLISVCCQSWTVSASNSQGLSSAKNYSELILEACIFGTFRDORLVQD
VOTFIENGHDCANIVMENTKREDHYRICAVKFCLLDGSNMHKLFIEDLAIKLL
```

CDS

```
DKDELVSCKRYYVNSLOHRVKNRVQWTLVLPPRLDQNFNGIIDIIRIFOAGFTNNO
ASIKYFIWIIILILHKEPQFLPKFWDVSGENLKTSLCTPLAVLSHLDIITQINIP
EKKDLKQALIVLWCFNFRFVRLYALVALKKLWTVCKVLSVEEDALTPVIESIL
HVQESHGAGNKNWQRIQEHFFFAFTHPLKDYCLETIIFYILPRLSGLIEDWITID
KFRFTDVPPLAAGFQWISQTSKLKPGDWSQDICTNLVNEADNQAEWTDVQKKIIP
WNSRVSDLLELLFQDRAARLGKISRLIVASLIDKPTNLGGLCRTCEVFGASVLVY
GSLQICSDKQFHLVSAGQWLPVLEVKPPQLIDYLOQKKTEGYTIIGVEQAKSLDL
GTCFPEKSLLLGNREGIPANLIQQLDVCVEIPQGGIIRSLNVHVSALLIWEYTR
QQLUSHGDTKP"
Join(complement (2360..2472),
complement (AL355472.19:90370..90519),
complement (AL355472.19:82690..82817))
/locus_tag="RP5-827C21.5-004"
Join(complement (2360..2472),
complement (AL355472.19:90370..90519),
complement (AL355472.19:82690..82817))
/locus_tag="RP5-827C21.5-004"
/product="TAR (HIV) RNA binding protein 1"
/notes="match: ESTs: Em:AA836178.1"
misc_feature
104632
/notes="Clone_right_end: RP4-792D7"
ORIGIN
Query Match 87.0%; Score 17.4; DB 8; Length 104632;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 22543 AAGTAAATGCAGACAGG 22561
|||||
RESULT 59
CR543861_32/c
WPCOMMENT
Sequence split into 36 fragments LOCUS CR543861 Accession CR543861
Fragment Name Begin End
CR543861_00 1 110000
CR543861_01 100001 210000
CR543861_02 200001 310000
CR543861_03 300001 410000
CR543861_04 400001 510000
CR543861_05 500001 610000
CR543861_06 600001 710000
CR543861_07 700001 810000
CR543861_08 800001 910000
CR543861_09 900001 1010000
CR543861_10 1000001 1110000
CR543861_11 1100001 1210000
CR543861_12 1200001 1310000
CR543861_13 1300001 1410000
CR543861_14 1400001 1510000
CR543861_15 1500001 1610000
CR543861_16 1600001 1710000
CR543861_17 1700001 1810000
CR543861_18 1800001 1910000
CR543861_19 1900001 2010000
CR543861_20 2000001 2110000
CR543861_21 2100001 2210000
CR543861_22 2200001 2310000
CR543861_23 2300001 2410000
CR543861_24 2400001 2510000
CR543861_25 2500001 2610000
CR543861_26 2600001 2710000
CR543861_27 2700001 2810000
CR543861_28 2800001 2910000
CR543861_29 2900001 3010000
CR543861_30 3000001 3110000
CR543861_31 3100001 3210000
CR543861_32 3200001 3310000
CR543861_33 3300001 3410000
CR543861_34 3400001 3510000
CR543861_35 3500001 3598621
```

Continuation (33 of 36) of CR543861 from base 3200001 (CR543861 Acinetobacter sp. ADP1 d

Query Match 87.0%; Score 17.4; DB 1; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 38781 CAAGTAAATGCAGAAACTG 38763

RESULT 60  
CR848719 2  
WPCOMMENT

Sequence split into 4 fragments LOCUS CR848719 Accession CR848719

Fragment Name	Begin	End
CR848719_0	1	110000
CR848719_1	100001	210000
CR848719_2	200001	310000
CR848719_3	300001	353035

Continuation (3 of 4) of CR848719 from base 200001 (CR848719 Danio rerio clone DKEY-269D

Query Match 87.0%; Score 17.4; DB 14; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
|||||  
DB 18237 AAGAAATGCAGAAACAGG 18255

Search completed: January 27, 2006, 23:14:09  
Job time : 793.944 secs

CR848719 2001 217

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 215.011 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23  
Sequence: 1 tgaagagcagtagaagcttag 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 60 summaries

Database : N Geneseq 21.1

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*
- 14: Geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	14	Adz75861 Group B s
2	23	100.0	1731	6	Abn68276 Streptoco
3	23	100.0	1734	13	Adv85417 Streptoco
4	23	100.0	29072	13	Adv87723 Streptoco
5	23	100.0	29072	13	Adv78976 Streptoco
6	23	100.0	110000	6	Continuation (8 of
7	23	100.0	110000	13	Continuation (9 of
8	19.8	86.1	1378	10	Adv68798 Streptoco
9	19.8	86.1	1378	12	Adj62002 Group A s
10	19.8	86.1	1379	10	Adv68800 Streptoco
11	19.8	86.1	1379	10	Adv68797 Streptoco
12	19.8	86.1	1379	10	Adv68791 Streptoco
13	19.8	86.1	1379	12	Adj62001 Group A s
14	19.8	86.1	1379	12	Adj62004 Group A s
15	19.8	86.1	1379	12	Adj61995 Group A s
16	19.8	86.1	1384	10	Adv68792 Streptoco
17	19.8	86.1	1384	10	Adv68794 Streptoco
18	19.8	86.1	1384	12	Adj61998 Group A s
19	19.8	86.1	1384	12	Adj61996 Group A s

20	19.8	86.1	1385	10	ADD68793	Ad68793 Streptoco
21	19.8	86.1	1385	12	ADJ61997	Adj61997 Group A s
22	19.8	86.1	1390	12	ADJ68795	Ad68795 Streptoco
23	19.8	86.1	1390	12	ADJ61999	Adj61999 Group A s
24	19.8	86.1	1393	10	ADJ68799	Ad68799 Streptoco
25	19.8	86.1	1393	12	ADD62003	Ad62003 Group A s
26	19.8	86.1	1447	10	ADD68801	Ad68801 Streptoco
27	19.8	86.1	1447	12	ADJ62005	Adj62005 Group A s
28	19.8	86.1	1731	8	ABN68277	Abn68277 Streptoco
29	19.8	86.1	1734	6	ACA50572	ACA50572 Prokaryot
30	19.8	86.1	1803	10	ADD68806	Ad68806 Streptoco
31	19.8	86.1	1803	12	ADJ61993	Adj61993 Group A s
32	18.8	81.7	89210	14	ADZ13911	Adz13911 Human can
33	18.2	79.1	720	9	ADA14456	Ada14456 Mouse spe
34	18.2	79.1	15766	10	ADE56804	Ade56804 Rat gene
35	18.2	79.1	110000	2	AAV21209_01	Continuation (2 of
36	18.2	79.1	160820	8	ABQ76673	Abq76673 Androgen
37	18	78.3	1423	10	ADC59504	Adc59504 Polypepti
38	17.8	77.4	1714	13	ADX49264	Adx49264 Plant ful
39	17.8	77.4	2533	11	ADM02795	Adm02795 Human cdn
40	17.8	77.4	6805	4	ABL13110	Ab113110 Drosoephil
41	17.8	77.4	9722	5	ABA15166	Ab15166 Human ner
42	17.8	77.4	14855	5	ABA15164	Ab15164 Human ner
43	17.8	77.4	22028	6	ABA93286	Ab93286 Human ace
44	17.8	77.4	22028	6	ABA93288	Ab93288 Human ace
45	17.8	77.4	104399	13	ABD33148	Abd33148 Murine ca
46	17.8	77.4	110000	2	AAV21209_12	Continuation (13 o
47	17.8	77.4	110000	2	AAV21209_13	Continuation (14 o
48	17.4	75.7	21	10	ADD68802	Ad68802 PCR prime
49	17.4	75.7	21	12	ADJ61989	Adj61989 Prol gene
50	17.4	75.7	309	4	AAK53660	Aak53660 Murine re
51	17.4	75.7	322	6	ABK63158	Abk63158 Rat seque
52	17.4	75.7	2151	14	ADZ62746	Adz62746 Murine Cd
53	17.4	75.7	7613	4	AAK65192	Aak65192 Human imm
54	17.4	75.7	24789	4	ABL28640	Ab128640 Drosoephil
55	17.2	74.8	519	6	ABK62529	Abk62529 Rat seque
56	17.2	74.8	519	10	ADB56078	Adb56078 Toxicity-
57	17.2	74.8	519	10	ADB50614	Adb50614 Primary r
58	17.2	74.8	670	6	ABL90230	Ab190230 Human pol
59	17.2	74.8	886	2	AAK97673	Aak97673 Extended
60	17.2	74.8	886	12	ADP18940	Adp18940 Human sec

ALIGNMENTS

RESULT 1

ADZ75861

ID ADZ75861 standard; DNA; 23 BP.

XX AC ADZ75861;

XX DT 28-JUL-2005 (first entry)

XX Group B streptococcus phosphotransferase (pts) PCR primer, SEQ ID NO:1.

XX Microorganism detection; fluorescence; diagnosis;  
streptococcus infection; infection; gynecology and obstetrics;  
phosphotransferase; PCR; primer; ss.

XX Streptococcus sp. 'group B'.

XX US2005106578-A1.

XX 19-MAY-2005.

XX 18-NOV-2003; 2003US-00716005.

XX 18-NOV-2003; 2003US-00716005.

XX (UHLJ/) UHL J R.

XX (COCK/) COCKERILL F R.

XX (AICH/) AICHINGER C.

```
PA (REIS/) REISER A.
XX
PI Uhl JR, Cockerill FR, Aichinger C, Reiser A;
XX WPI; 2005-371550/38.
XX
PT Detecting group B streptococcus, comprises amplifying a sample with pts
PT primers, hybridizing the sample with fluorescently labeled pts probes,
PT and detecting the presence of fluorescence resonance energy transfer.
XX
XX Claim 2; SEQ ID NO 1; 13pp; English.
XX
XX The invention relates to a real-time PCR-based method of detecting the
XX presence or absence of group B streptococcus (GBS) bacterial pathogens in
XX a biological sample from an individual. The method comprises
XX amplification of a conserved region of the phosphotransferase (pts) gene
XX of GBS (especially using PCR primers AD275861-AD275862), detection of the
XX amplification product with a pair of fluorescently labeled
XX phosphotransferase probes (especially AD275863-AD275864), and detection
XX of the presence or absence of fluorescence resonance energy transfer
XX (FRET), where presence of FRET is indicative of the presence of GBS in
XX the sample. GBS infection is a leading cause of neonatal morbidity and
XX mortality, with infection occurring during childbirth. Currently, it is
XX recommended that women are screened for GBS during week 35-37 of
XX gestation by a culture-based method which may take up to 72 hours for a
XX result. However, many women first present at healthcare facilities at the
XX time of labor, and in addition, GBS infection can be transient, so that a
XX woman free of GBS at the time of screening may not be free of GBS when
XX she is due to give birth. The method of the invention provides a real-
XX time assay for the detection of group B streptococcus in a sample, is
XX more sensitive and specific than prior art non-culture based methods, and
XX can thus be implemented for routine diagnosis of the presence of group B
XX streptococcus. Sequences AD275861-AD275862 represent GBS
XX phosphotransferase PCR primers specifically claimed for use in the method
XX of the invention.
SQ Sequence 23 BP; 9 A; 2 C; 8 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 14; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAAGCGCAGTAGAAGCTTAG 23
DB 1 TGAGAAGCGCAGTAGAAGCTTAG 23

RESULT 2
ID ABN68276
XX ABN68276 standard; DNA; 1731 BP.
XX
AC ABN68276;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 4465.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; bacterial infection; ds.
XX
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
```

```
PA (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI; 2002-352536/38.
DR P-P8DB; ABP27645.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3607; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins
XX
XX Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAAGCGCAGTAGAAGCTTAG 23
DB 177 TGAGAAGCGCAGTAGAAGCTTAG 199

RESULT 3
ID ADV85417
XX ADV85417 standard; DNA; 1734 BP.
XX
AC ADV85417;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae DNA sequence, SEQ ID 6558.
XX
XX Antibacterial; vaccine; bacterial infection; ds.
XX
OS Streptococcus agalactiae.
XX
XX WO200292818-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-IB003059.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
```

XX Genomic nucleotide sequences encoding polypeptides of *Streptococcus*  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
PT and identification of therapeutic targets.  
XX  
PS Claim 4; SEQ ID NO 6558; 439pp; French.  
XX  
XX The present invention relates to novel *Streptococcus agalactiae*  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transporters, cell membrane proteins and  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 23; DB 13; Length 1734;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
DB 177 TGAGAAGGCAGTAGAAGCTTAG 199  
RESULT 4  
ADV87723  
ID ADV87723 standard; DNA; 29072 BP.  
AC ADV87723;  
XX  
XX 24-FEB-2005 (first entry)  
DE  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
XX Antibacterial; Vaccine; bacterial infection; ds.  
XX Streptococcus agalactiae.  
OS  
XX FR2824074-A1.  
PN  
XX 31-OCT-2002.  
PD  
XX  
XX 26-APR-2001; 2001FR-00005642.  
PF  
XX  
XX 26-APR-2001; 2001FR-00005642.  
PR  
XX  
XX (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
PI  
XX WPI; 2004-101891/11.  
DR  
XX  
XX Genomic nucleotide sequences encoding polypeptides of *Streptococcus*  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
PT and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 2687pp; French.  
XX  
XX The present invention relates to novel *Streptococcus agalactiae*

CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of *S.*  
CC agalactiae involved in the synthesis of amino acids, cell membranes,  
CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
CC phospholipid metabolism, nucleotide metabolism including purines,  
CC pyrimidines and/or nucleosides, regulatory functions, replication,  
CC transcription, translation, protein transport, adaptation to atypical  
CC conditions, sensitivity to medicines and/or analogues, functions related  
CC to transporters, cell membrane proteins and cellular machinery. (I) are  
CC useful for the detection and/or amplification of nucleic acids.  
CC Pharmaceutical composition comprising (I) or (II) are useful for  
CC treatment of a bacterial *S. agalactiae* infection. Note: WO20022818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO20022818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX  
XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 23; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
DB 12072 TGAGAAGGCAGTAGAAGCTTAG 12094  
RESULT 5  
ADV78976  
ID ADV78976 standard; DNA; 29072 BP.  
XX  
XX ADV78976;  
AC  
XX  
XX 24-FEB-2005 (first entry)  
DT  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
DE  
XX  
XX Antibacterial; vaccine; bacterial infection; ds.  
KW  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO20022818-A2.  
XX  
XX 21-NOV-2002.  
PD  
XX  
XX 26-APR-2002; 2002WO-IB003059.  
PF  
XX  
XX 26-APR-2001; 2001FR-00005642.  
PR  
XX  
XX (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
PI  
XX WPI; 2004-101891/11.  
DR  
XX  
XX Genomic nucleotide sequences encoding polypeptides of *Streptococcus*  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
PT and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 439pp; French.  
XX  
XX The present invention relates to novel *Streptococcus agalactiae*  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines

CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 13; Length 29072;

Best Local Similarity 100.0%; Pred. No. 2.9; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 12072 TGAGAAGGCAGTAGAAGCTTAG 12094

#### RESULT 6

ABN71527\_07  
Continuation (8 of 22) of ABN71527 from base 700001 (*Streptococcus polynucleotide* SEQ ID  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	1000001	1110000
WP ABN71527_11	1100001	1210000
WP ABN71527_12	1200001	1310000
WP ABN71527_13	1300001	1410000
WP ABN71527_14	1400001	1510000
WP ABN71527_15	1500001	1610000
WP ABN71527_16	1600001	1710000
WP ABN71527_17	1700001	1810000
WP ABN71527_18	1800001	1910000
WP ABN71527_19	1900001	2010000
WP ABN71527_20	2000001	2110000
WP ABN71527_21	2100001	2155561

Query Match 100.0%; Score 23; DB 6; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 90942 TGAGAAGGCAGTAGAAGCTTAG 90964

#### RESULT 7

ADV81204\_08  
Continuation (9 of 23) of ADV81204 from base 800001 (*Streptococcus agalactiae* complete g  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession ADV81204

Fragment Name	Begin	End
WP ADV81204_00	1	110000
WP ADV81204_01	100001	210000
WP ADV81204_02	200001	310000
WP ADV81204_03	300001	410000
WP ADV81204_04	400001	510000
WP ADV81204_05	500001	610000
WP ADV81204_06	600001	710000
WP ADV81204_07	700001	810000
WP ADV81204_08	800001	910000
WP ADV81204_09	900001	1010000
WP ADV81204_10	1000001	1110000

WP	ADV81204_11	1100001	1210000
WP	ADV81204_12	1200001	1310000
WP	ADV81204_13	1300001	1410000
WP	ADV81204_14	1400001	1510000
WP	ADV81204_15	1500001	1610000
WP	ADV81204_16	1600001	1710000
WP	ADV81204_17	1700001	1810000
WP	ADV81204_18	1800001	1910000
WP	ADV81204_19	1900001	2010000
WP	ADV81204_20	2000001	2110000
WP	ADV81204_21	2100001	2210000
WP	ADV81204_22	2200001	2217924

Query Match 100.0%; Score 23; DB 13; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 71277 TGAGAAGGCAGTAGAAGCTTAG 71299

#### RESULT 8

ADD68798

ID ADD68798 standard; DNA; 1378 BP.

XX ADD68798;

AC ADD68798;

XX 15-JAN-2004 (first entry)

DT Streptococcus sp. 'group A' *ptxI* DNA - SEQ ID 14.

DE beta-haemolytic Group A *Streptococcus*; GAS;

KW fluorescence resonance energy transfer; FRET; *ptxI*; *ds*.

XX Streptococcus sp. 'group A'.

OS US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

PT Detecting presence/absence of group A *Streptococcus* in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 14; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A *Streptococcus* (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescently-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the *Streptococcus* sp. 'group A' *ptxI* DNA (SEQ ID 14)  
CC of the invention.

XX Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;



DE	Streptococcus sp. 'group A' ptaI DNA - SEQ ID 16.
XX	
XX	beta-haemolytic Group A Streptococcus; GAS;
KW	fluorescence resonance energy transfer; FRET; ptaI; de.
XX	
XX	Streptococcus sp. 'group A'.
OS	
XX	US6593093-B1.
PN	
XX	15-JUL-2003.
PD	
XX	
PF	20-FEB-2002; 2002US-00081923.
XX	
XX	20-FEB-2002; 2002US-00081923.
PR	
XX	(MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
PA	
XX	Uhl JR, Cockerill FR;
PI	
XX	WPI; 2003-828265/77.
DR	
XX	Detecting presence/absence of group A Streptococcus in sample by
PT	performing cycling step comprising amplifying, hybridizing DNA from
PT	sample by fluorescent probe, detecting for fluorescent resonance energy
PT	transfer.
XX	
XX	Example 2; SEQ ID NO 16; 35pp; English.
PS	
XX	The invention relates to a novel method for detecting the presence or
XX	absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC	isolated from an individual by performing more than one cycling step. The
CC	method involves an amplifying step comprising contacting the sample with
CC	GAS primers, a subsequent hybridizing step comprising contacting the
CC	amplified product with fluorescent-labelled probes and finally, detecting
CC	fluorescence resonance energy transfer (FRET) where the presence or
CC	absence of FRET indicates the presence or absence of GAS in the sample.
CC	The method of the invention may be useful for detecting the presence or
CC	absence of GAS in a biological sample from an individual. The current
CC	sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 16)
CC	of the invention.
XX	
XX	Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;
SQ	
	Query Match 86.1%; Score 19.8; DB 10; Length 1379;
	Best Local Similarity 91.3%; Pred. No. 58;
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGAGTAGAAGCTTAG 23
Db	10 TGAAATGTCAGTAGAAGCTTAG 32
RESULT 11	
ADD68797	ID ADD68797 standard; DNA; 1379 BP.
XX	
AC	ADD68797;
XX	
XX	15-JAN-2004 (first entry)
DT	
XX	Streptococcus sp. 'group A' ptaI DNA - SEQ ID 13.
DE	
XX	beta-haemolytic Group A Streptococcus; GAS;
KW	fluorescence resonance energy transfer; FRET; ptaI; de.
XX	
OS	Streptococcus sp. 'group A'.
XX	
XX	US6593093-B1.
PN	
XX	15-JUL-2003.
PD	
XX	20-FEB-2002; 2002US-00081923.
PF	
XX	

```
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 13; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 13)
XX of the invention.
XX
XX Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1379;
XX Best Local Similarity 91.3%; Pred. No. 58;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAGCTTAG 31
XX
XX RESULT 12
XX ADD68791
XX ID ADD68791 standard; DNA; 1379 BP.
XX
XX AC ADD68791;
XX
XX DT 15-JAN-2004 (first entry)
XX
XX DE Streptococcus sp. 'group A' ptsl DNA - SEQ ID 7.
XX
XX KW beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptsl; ds.
XX
XX OS Streptococcus sp. 'group A'.
XX
XX PN US6593093-B1.
XX
XX PD 15-JUL-2003.
XX
XX PF 20-FEB-2002; 2002US-00081923.
XX
XX PR 20-FEB-2002; 2002US-00081923.
XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX PI Uhl JR, Cockerill FR;
XX
XX XX WPI; 2003-828265/77.
XX
XX PT Detecting presence/absence of group A Streptococcus in sample by
XX performing cycling step comprising amplifying, hybridizing DNA from
XX sample by fluorescent probe, detecting for fluorescent resonance energy
XX transfer.
XX
XX Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1379;
XX Best Local Similarity 91.3%; Pred. No. 58;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAGCTTAG 31
XX
XX RESULT 13
XX ADJ62001
XX ID ADJ62001 standard; DNA; 1379 BP.
XX
XX AC ADJ62001;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE Group A streptococcus ptsl gene #8.
XX
XX KW Group A Streptococcus; GAS; gene; ds.
XX
XX OS Streptococcus sp.
XX
XX PN US2004014118-A1.
XX
XX PD 22-JAN-2004.
XX
XX PF 19-JUN-2003; 2003US-00465205.
XX
XX PR 20-FEB-2002; 2002US-00081923.
XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX PI Uhl JR, Cockerill FR;
XX
XX DR WPI; 2004-224229/21.
XX
XX PT Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsl primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX PS Disclosure; SEQ ID NO 13; 36pp; English.
XX
XX CC The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsl primers and probes to produce ptsl
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsl gene.
XX
XX Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;
```

Query Match 86.1%; Score 19.8; DB 12; Length 1379;  
Best Local Similarity 91.3%; Pred. No. 58;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
DB 9 TGAATAATGCAGTAGAAGCTTAG 31

RESULT 14  
ADJ62004  
ID ADJ62004 standard; DNA; 1379 BP.

XX AC ADJ62004;  
XX DT 06-MAY-2004 (first entry)  
XX DE Group A streptococcus PtsI gene #11.  
XX KW Group A Streptococcus; GAS; gene; ds.  
XX OS Streptococcus sp.  
XX PN US2004014118-A1.  
XX PD 22-JAN-2004.

XX PF 19-JUN-2003; 2003US-00465205.  
XX PR 20-FEB-2002; 2002US-00081923.  
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX PI Uhl JR, Cockerill FR;  
XX DR WPI; 2004-224229/21.

XX PT Detecting Group A Streptococcus in biological sample, involves contacting sample with pair of ptsI primers and probes and detecting fluorescence  
XX PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX PS Disclosure; SEQ ID NO 16; 36pp; English.

XX CC The invention relates to a method of detecting the presence or absence of Group A Streptococcus (GAS) in biological sample from individual, by contacting sample with pair of ptsI primers and probes to produce ptsI amplification product, where first, second probes are labelled with donor and acceptor fluorescent moiety, and detecting presence or absence of fluorescence resonance energy transfer, which indicates the presence or absence of GAS in biological sample. The method is useful for detecting the presence or absence of Group A Streptococcus in a biological sample such as throat swabs, tissues and bodily fluids from an individual. The present sequence represents a group A streptococcus ptsI gene.

XX SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 12; Length 1379;  
Best Local Similarity 91.3%; Pred. No. 58;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
DB 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 15  
ADJ61995  
ID ADJ61995 standard; DNA; 1379 BP.

XX AC ADJ61995;  
XX DT 06-MAY-2004 (first entry)

DE Group A streptococcus PtsI gene #2.  
XX KW Group A Streptococcus; GAS; gene; ds.  
XX OS Streptococcus sp.  
XX PN US2004014118-A1.  
XX PD 22-JAN-2004.  
XX PF 19-JUN-2003; 2003US-00465205.  
XX PR 20-FEB-2002; 2002US-00081923.  
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Uhl JR, Cockerill FR;  
XX DR WPI; 2004-224229/21.

XX PT Detecting Group A Streptococcus in biological sample, involves contacting sample with pair of ptsI primers and probes and detecting fluorescence  
XX PT resonance energy transfer indicating Group A Streptococcus in sample.

XX PS Disclosure; SEQ ID NO 7; 36pp; English.

XX CC The invention relates to a method of detecting the presence or absence of Group A Streptococcus (GAS) in biological sample from individual, by contacting sample with pair of ptsI primers and probes to produce ptsI amplification product, where first, second probes are labelled with donor and acceptor fluorescent moiety, and detecting presence or absence of fluorescence resonance energy transfer, which indicates the presence or absence of GAS in biological sample. The method is useful for detecting the presence or absence of Group A Streptococcus in a biological sample such as throat swabs, tissues and bodily fluids from an individual. The present sequence represents a group A streptococcus ptsI gene.

XX SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 12; Length 1379;  
Best Local Similarity 91.3%; Pred. No. 58;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
DB 9 TGAATAATGCAGTAGAAGCTTAG 31

RESULT 16  
ADD68792  
ID ADD68792 standard; DNA; 1384 BP.

XX AC ADD68792;

XX DT 15-JAN-2004 (first entry)

XX DE Streptococcus sp. 'group A' ptsI DNA - SEQ ID 8.  
XX KW beta-haemolytic Group A Streptococcus; GAS;  
XX KW fluorescence resonance energy transfer; FRET; ptsI; ds.

XX OS Streptococcus sp. 'group A'.

XX PN US6593093-B1.

XX PD 15-JUL-2003.

XX PF 20-FEB-2002; 2002US-00081923.

XX PR 20-FEB-2002; 2002US-00081923.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

```

PI Uhl JR, Cockerill FR;
XX WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 8; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 8)
XX
XX SQ Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 10 TGAATAATGCAGTAGAAGCTTAG 32
XX
XX RESULT 17
XX ADD68794
XX ID ADD68794 standard; DNA; 1384 BP.
XX AC ADD68794;
XX AC
XX 15-JAN-2004 (first entry)
XX DT
XX DE Streptococcus sp. 'group A' ptsl DNA - SEQ ID 10.
XX KW beta-haemolytic Group A Streptococcus; GAS;
XX KW fluorescence resonance energy transfer; FRET; ptsl; da.
XX OS Streptococcus sp. 'group A'.
XX XX
XX PN US5593093-B1.
XX XX
XX PD 15-JUL-2003.
XX XX
XX PF 20-FEB-2002; 2002US-00081923.
XX XX
XX PR 20-FEB-2002; 2002US-00081923.
XX XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;
XX XX
XX DR WPI; 2003-828265/77.
XX XX
XX PT Detecting presence/absence of group A Streptococcus in sample by
XX PT performing cycling step comprising amplifying, hybridizing DNA from
XX PT sample by fluorescent probe, detecting for fluorescent resonance energy
XX PT transfer.
XX
XX Example 2; SEQ ID NO 10; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 8)
XX
XX SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 9 TGAATAATGCAGTAGAAGCTTAG 31
XX
XX RESULT 18
XX ADJ61998
XX ID ADJ61998 standard; DNA; 1384 BP.
XX AC ADJ61998;
XX AC
XX 06-MAY-2004 (first entry)
XX DT
XX DE Group A streptococcus ptsl gene #5.
XX XX
XX KW Group A Streptococcus; GAS; gene; da.
XX XX
XX OS Streptococcus sp.
XX XX
XX PN US2004014118-A1.
XX XX
XX PD 22-JAN-2004.
XX XX
XX PF 19-JUN-2003; 2003US-00465205.
XX XX
XX PR 20-FEB-2002; 2002US-00081923.
XX XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;
XX XX
XX DR WPI; 2004-224229/21.
XX XX
XX PT Detecting Group A Streptococcus in biological sample, involves contacting
XX PT sample with pair of ptsl primers and probes and detecting fluorescence
XX PT resonance energy transfer indicating Group A Streptococcus in sample.
XX XX
XX PS Disclosure; SEQ ID NO 10; 36pp; English.
XX XX
XX CC The invention relates to a method of detecting the presence or absence of
XX CC Group A Streptococcus (GAS) in biological sample from individual, by
XX CC contacting sample with pair of ptsl primers and probes to produce ptsl
XX CC amplification product, where first, second probes are labelled with donor
XX CC and acceptor fluorescent moiety, and detecting presence or absence of
XX CC fluorescence resonance energy transfer, which indicates the presence or
XX CC absence of GAS in biological sample. The method is useful for detecting
XX CC the presence or absence of Group A Streptococcus in a biological sample
XX CC such as throat swabs, tissues and bodily fluids from an individual. The
XX CC present sequence represents a group A streptococcus ptsl gene.
XX
XX SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 12; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
 DB 9 TGAATAATGCAGTAGAAGCTTAG 31

## RESULT 19

ADJ61996  
 ID ADJ61996 standard; DNA; 1384 BP.  
 AC ADJ61996;  
 XX 06-MAY-2004 (first entry)  
 XX Group A streptococcus PtsI gene #3.  
 XX Group A streptococcus; GAS; gene; ds.  
 XX Streptococcus sp.  
 XX US2004014118-A1.  
 XX 22-JAN-2004.

XX 19-JUN-2003; 2003US-00465205.  
 XX 20-FEB-2002; 2002US-00081923.  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX Uhl JR, Cockerill FR;

XX WPI; 2004-224229/21.

XX Detecting Group A Streptococcus in biological sample, involves contacting sample with pair of ptsI primers and probes and detecting fluorescence resonance energy transfer indicating Group A Streptococcus in sample.

XX Disclosure; SEQ ID NO 8; 36pp; English.

XX The invention relates to a method of detecting the presence or absence of Group A Streptococcus (GAS) in biological sample from individual, by contacting sample with pair of ptsI primers and probes to produce ptsI amplification product, where first, second probes are labelled with donor and acceptor fluorescent moiety, and detecting presence or absence of fluorescence resonance energy transfer, which indicates the presence or absence of GAS in biological sample. The method is useful for detecting the presence or absence of Group A Streptococcus in a biological sample such as throat swabs, tissues and bodily fluids from an individual. The present sequence represents a group A streptococcus ptsI gene.

XX Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 12; Length 1384;  
 Best Local Similarity 91.3%; Pred. No. 59;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
 DB 10 TGAATAATGCAGTAGAAGCTTAG 32

## RESULT 20

ADJ68793  
 ID ADD68793 standard; DNA; 1385 BP.

XX ADD68793;

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptsI DNA - SEQ ID 9.

XX beta-haemolytic Group A Streptococcus; GAS;  
 XX fluorescence resonance energy transfer; FRET; ptsI; ds.

XX Streptococcus sp. 'group A'.  
 OS US6593093-B1.  
 PN 15-JUL-2003.  
 XX 20-FEB-2002; 2002US-00081923.  
 XX 20-FEB-2002; 2002US-00081923.  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX Uhl JR, Cockerill FR;  
 XX WPI; 2003-828265/77.

XX Detecting presence/absence of group A Streptococcus in sample by performing cycling step comprising amplifying, hybridizing DNA from sample by fluorescent probe, detecting for fluorescent resonance energy transfer.

XX Example 2; SEQ ID NO 9; 35pp; English.

XX The invention relates to a novel method for detecting the presence or absence of beta-haemolytic Group A Streptococcus (GAS) in a sample isolated from an individual by performing more than one cycling step. The method involves an amplifying step comprising contacting the sample with GAS primers, a subsequent hybridising step comprising contacting the amplified product with fluorescent-labelled probes and finally, detecting fluorescence resonance energy transfer (FRET) where the presence or absence of FRET indicates the presence or absence of GAS in the sample. The method of the invention may be useful for detecting the presence or absence of GAS in a biological sample from an individual. The current sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 9) of the invention.

XX Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 10; Length 1385;  
 Best Local Similarity 91.3%; Pred. No. 59;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
 DB 10 TGAATAATGCAGTAGAAGCTTAG 32

## RESULT 21

ADJ61997  
 ID ADJ61997 standard; DNA; 1385 BP.

XX ADJ61997;

XX 06-MAY-2004 (first entry)

XX Group A streptococcus PtsI gene #4.

XX Group A Streptococcus; GAS; gene; ds.

XX Streptococcus sp.

XX US2004014118-A1.

XX 22-JAN-2004.

XX 19-JUN-2003; 2003US-00465205.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2004-224229/21.  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX Disclosure; SEQ ID NO 9; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsI gene.  
XX  
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 12; Length 1385;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 10 TGAATAATGCAGTAGAAGCTTAG 32  
RESULT 22  
ADD68795  
ID ADD68795 standard; DNA; 1390 BP.  
XX AC  
XX ADD68795;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Streptococcus sp. 'group A' ptsI DNA - SEQ ID 11.  
XX  
KW beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptsI; ds.  
XX  
OS Streptococcus sp. 'group A'.  
XX  
XX US6593093-B1.  
XX  
XX 15-JUL-2003.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
XX Uhl JR, Cockerill PR;  
XX  
XX WPI; 2003-828265/77.  
XX  
XX Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
XX Example 2; SEQ ID NO 11; 35pp; English.  
XX  
XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridizing step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting

CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 11)  
CC of the invention.  
XX  
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 10; Length 1390;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 24 TGAATAATGCAGTAGAAGCTTAG 46  
RESULT 23  
ADJ61999  
ID ADJ61999 standard; DNA; 1390 BP.  
XX AC  
XX ADJ61999;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
XX Group A streptococcus ptsI gene #6.  
DE  
XX Group A Streptococcus; GAS; gene; ds.  
KW  
XX Streptococcus sp.  
OS  
XX US2004014118-A1.  
PN  
XX 22-JAN-2004.  
PD  
XX  
XX 19-JUN-2003; 2003US-00465205.  
PF  
XX  
XX 20-FEB-2002; 2002US-00081923.  
PR  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
PA  
XX Uhl JR, Cockerill PR;  
PI  
XX  
XX WPI; 2004-224229/21.  
DR  
XX  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
XX Disclosure; SEQ ID NO 11; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsI gene.  
XX  
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 12; Length 1390;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 24 TGAATAATGCAGTAGAAGCTTAG 46

```

RESULT 24
ADD68799
ID ADD68799 standard; DNA; 1393 BP.
XX
XX AC ADD68799;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptst DNA - SEQ ID 15.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptstI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
XX performing cycling step comprising amplifying, hybridizing DNA from
XX sample by fluorescent probe, detecting for fluorescent resonance energy
XX transfer.
XX
XX Example 2; SEQ ID NO 15; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridizing step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptstI DNA (SEQ ID 15)
XX of the invention.
XX
XX SQ Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1393;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

The invention relates to a novel method for detecting the presence or absence of
Group A Streptococcus (GAS) in a biological sample from individual, by
contacting sample with pair of ptstI primers and probes to produce ptstI
amplification product, where first, second probes are labelled with donor
and acceptor fluorescent moiety, and detecting presence or absence of
fluorescence resonance energy transfer, which indicates the presence or
absence of GAS in biological sample. The method is useful for detecting
the presence or absence of Group A Streptococcus in a biological sample
such as throat swabs, tissues and bodily fluids from an individual. The
current sequence represents a group A streptococcus ptstI gene.

XX SQ Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 12; Length 1393;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGAGAGGCGAGTAGAAGCTTAG 23
DB 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 26
ADD68801
ID ADD68801 standard; DNA; 1447 BP.
XX
XX AC ADD68801;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptstI DNA - SEQ ID 17.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptstI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Group A streptococcus ptstI gene #10.
XX
XX Group A Streptococcus; GAS; gene; ds.

```

```

OS Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptstI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 15; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptstI primers and probes to produce ptstI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptstI gene.
XX
XX SQ Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 12; Length 1393;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGAGAGGCGAGTAGAAGCTTAG 23
DB 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 26
ADD68801
ID ADD68801 standard; DNA; 1447 BP.
XX
XX AC ADD68801;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptstI DNA - SEQ ID 17.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptstI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX

```





SQ Sequence 1731 BP; 517 A; 358 C; 389 G; 467 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 6; Length 1731;  
Best Local Similarity 91.3%; Pred. No. 60;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAGGCGAGTGAAGAGCTTAG 23  
Db 177 TGAATAATGCAGTGAAGAGCTTAG 199  
RESULT 29  
ACAS0572  
ID ACAS0572 standard; DNA; 1734 BP.  
XX AC ACAS0572;  
XX DT 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #32229.  
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX KW drug design; gene.  
XX OS Streptococcus pyogenes.  
XX PN WO20027183-A2.  
XX PD 03-OCT-2002.  
XX PP 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU46702.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 38442; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1734 BP; 519 A; 358 C; 389 G; 468 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 8; Length 1734;

Best Local Similarity 91.3%; Pred. No. 60;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGAGCTTAG 23

Db 177 TGAATAATGCAGTGAAGAGCTTAG 199

RESULT 30

ADD68806

ID ADD68806 standard; DNA; 1803 BP.

XX AC ADD68806;

XX DT 15-JAN-2004 (first entry)

XX DE Streptococcus sp. 'group A' *ptxI* DNA - SEQ ID 5.

XX KW beta-haemolytic Group A Streptococcus; GAS;

XX KW fluorescence resonance energy transfer; FRET; *ptxI*; ds, gene.

XX OS Streptococcus sp. 'group A'.

XX FH Key Location/Qualifiers

XX CDS 1..1803

XX FT /\*tag= a

XX FT /partial

XX FT /product= "Streptococcus sp. 'group A' *ptxI* protein - SEQ

XX FT ID 6"

XX FT /note= "No stop codon; this CDS contains translation

XX FT exceptions"

XX PN US6593093-B1.

XX PD 15-JUL-2003.

XX PP 20-FEB-2002; 2002US-00081923.

XX PR 20-FEB-2002; 2002US-00081923.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

XX P-PSDB; ADD68790.

XX Detecting presence/absence of group A Streptococcus in sample by  
XX performing cycling step comprising amplifying, hybridizing DNA from  
XX sample by fluorescent probe, detecting for fluorescent resonance energy  
XX transfer.

XX Example 2; SEQ ID NO 5; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridising step comprising contacting the



CC capacity of individuals.  
XX  
SQ Sequence 720 BP; 171 A; 200 C; 171 G; 178 T; 0 U; 0 Other;  
Query Match 79.1%; Score 18.2; DB 9; Length 720;  
Best Local Similarity 87.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||  
DB 499 TGAGAAGGCAGTAGAAGCAGAG 477  
RESULT 34  
ADE56804  
ID ADE56804 standard; DNA; 15766 BP.  
XX  
AC ADE56804;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Rat gene J05677, SEQ ID NO 2659.  
XX  
Rat; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI: 2003-268312/26.  
GENBANK; J05677.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
The invention discloses a composition comprising two or more isolated rat  
or human polynucleotides or a polynucleotide which represents a fragment,  
derivative or allelic variation of the nucleic acid sequence. Also  
claimed are a vector comprising the novel polynucleotide, a host cell  
comprising the vector, a method for identifying a nucleotide sequence  
which is differentially regulated in an animal subjected to pain and a  
kit to perform the method, an array, a method for identifying an agent  
that increases or decreases the expression of the polynucleotide sequence  
that is differentially expressed in neuronal tissue of a first animal  
subjected to pain, a method for identifying a compound which regulates  
the expression of a polynucleotide sequence which is differentially  
expressed in an animal subjected to pain, a method for identifying a  
compound that regulates the activity of one or more of the  
polynucleotides, a method for producing a pharmaceutical composition, a  
method for identifying a compound or small molecule that regulates the  
activity in an animal of one or more of the polypeptides given in the  
specification, a method for identifying a compound useful in treating  
pain and a pharmaceutical composition comprising the one or more  
polypeptides or their antibodies. The polynucleotide or the compound that  
modulates its activity is useful for preparing a medicament for treating  
pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

QY 2 GAGAAGGCAGTAGAAGCTTAG 23  
|||||  
DB 4526 GAGAAGGAAGTAGAAGCCTAG 4505  
RESULT 33  
ADA14456/C  
ID ADA14456 standard; cDNA; 720 BP.  
XX  
AC ADA14456;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Mouse spermatogenesis related cDNA sequence SEQ ID NO:15.  
XX  
mouse; spermatogenesis; gene cluster; mutagenicity;  
KW reproductive toxicity; reproductive capacity; mutation;  
KW expression abnormality; human male sterility associated gene; scot-t;  
KW succinyl CoA:3-oxo acid CoA transferase; human male sterility; gene; ss.  
XX  
OS Mus musculus.  
XX  
PN WO2003068969-A1.  
XX  
PD 21-AUG-2003.  
XX  
PF 14-FEB-2003; 2003WO-JP001572.  
XX  
PR 14-FEB-2002; 2002JP-00036649.  
PR 27-DEC-2002; 2002JP-00381241.  
XX  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX  
PI Nishimune Y, Tanaka H, Nozaki M;  
XX  
DR WPI: 2003-671663/63.  
XX  
PT Mouse spermatogenesis gene cluster and human male sterility associated  
genes, useful for diagnosis of human male sterility and testing  
substances for reproductive toxicity.  
XX  
PS Claim 1; Page 70-71; 262pp; Japanese.  
XX  
The present invention describes a mouse spermatogenesis gene cluster  
containing 89 genes (see the cDNA sequences of ADA14442 to ADA14530).  
CC Also described: (1) a cDNA library containing cDNA encoding the gene  
CC cluster; (2) oligonucleotides of 10-99 bases containing partial sequences  
CC of genes of the cluster; (3) microarrays containing these  
CC oligonucleotides; (4) primer sets for PCR amplification of cDNA or  
CC genomic DNA for genes of the cluster; (5) polypeptides encoded by the  
CC genes in the cluster; (6) antibodies to these polypeptides; (7) a method  
CC for testing the mutagenicity and reproductive toxicity of a test  
CC substance, and assessment of the reproductive capacity of a test  
CC individual, by analysis of mutation and expression abnormalities of genes  
CC in the cluster; (8) polynucleotides which are mutations of the human male  
CC sterility associated gene scot-t (succinyl CoA:3-oxo acid CoA transferase  
CC gene) having one or more of the following specific mutations: T129C,  
CC T870G, C1071T, T1667C; (9) oligonucleotides containing partial sequences  
CC of human scot-t including one or more of the above mutations; (10) primer  
CC sets for PCR amplification of mRNA derived from the mutant scot-t gene;  
CC (11) polypeptides encoded by human scot-t gene and having one or more of  
CC the mutations Leu38Pro, Leu285Arg, Thr352Met; (12) polynucleotides which  
CC are mutations of the human male sterility associated gene protamine2,  
CC having C248T; (13) polypeptides encoded by this mutant protamine2 gene;  
CC (14) antibodies (including labelled antibodies) to these polypeptides;  
CC (15) a method for determining the presence or absence of these mutant  
CC polynucleotides in genomic DNA; (16) diagnosis of human male sterility  
CC using this method; (17) DNA probes containing sequences of these mutant  
CC polynucleotides; and (18) DNA chip containing sequences derived from  
CC these mutant polynucleotides. The methods of the present invention can be  
CC used in the diagnosis of human male sterility; testing the reproductive  
CC toxicity and mutagenicity of substances; and assessing the reproductive

```
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15766 BP; 3852 A; 4174 C; 3768 G; 3972 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 10; Length 15766;
Best Local Similarity 87.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 6272 TGTGAGGCGAGTAGAAGCATAG 6294
RESULT 35
AAV21209_01
Continuation (2 of 17) of AAV21209 from base 100001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name Begin End
WP AAV21209_00 1 110000
WP AAV21209_01 100001 210000
WP AAV21209_02 200001 310000
WP AAV21209_03 300001 410000
WP AAV21209_04 400001 510000
WP AAV21209_05 500001 610000
WP AAV21209_06 600001 710000
WP AAV21209_07 700001 810000
WP AAV21209_08 800001 910000
WP AAV21209_09 900001 1010000
WP AAV21209_10 1000001 1110000
WP AAV21209_11 1100001 1210000
WP AAV21209_12 1200001 1310000
WP AAV21209_13 1300001 1410000
WP AAV21209_14 1400001 1510000
WP AAV21209_15 1500001 1610000
WP AAV21209_16 1600001 1664976
Query Match 79.1%; Score 18.2; DB 2; Length 110000;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 69531 TGAAAAGGCATTAAAAAGCTTAG 69553
RESULT 36
ABQ76673
ID ABQ76673 standard; DNA; 160820 BP.
XX
AC ABQ76673;
XX
DT 26-MAR-2003 (first entry)
XX
DE Androgen receptor signalling pathway-associated DNA AB043547.
XX
KW Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;
KW signal transduction pathway; transforming growth factor-B; phosphatase;
KW tensin; cytoskeletal; antiproliferative; cellular proliferation; cancer;
KW AB043547; ds.
XX
OS Synthetic.
XX
PN WO200282081-A2.
XX
PD 17-OCT-2002.
XX
PF 05-APR-2002; 2002WO-US011086.
XX
```

```
XX 06-APR-2001; 2001US-0282266P.
PR 13-MAR-2002; 2002US-0385060P.
XX
XX (UYRP ) UNIV ROCHESTER.
XX Chang C;
XX
XX WPI; 2003-046871/04.
DR
XX
PT Modulating androgen receptor activity, by administering a compound that
PT modulates receptor activity, inhibits receptor-signal transduction
PT pathway/receptor-coactivator interaction or changes amount or receptor.
XX
XX Disclosure; Page 181-225; 302pp; English.
XX
CC This invention describes a novel method for modulating androgen receptor
CC activity or androgen receptor-mediated transactivation activity in a
CC cell. The method involves administering a compound which causes
CC modulation of the androgen receptors activity and the inhibition of
CC interaction between the receptor and a protein involved in a signal
CC transduction pathway. The compound also inhibits the interaction between
CC the androgen receptor and a protein selected from Smad3, Smad4, Akt,
CC transforming growth factor (TGF)-B and phosphatase and tensin homologues
CC deleted on chromosome 10 (PTEN) or their fragments. The compounds of the
CC invention have cytostatic and antiproliferative activity. The obtained
CC composition is useful for treating any disease, where uncontrolled
CC proliferation or cellular proliferation occurs such as cancer, e.g.
CC prostate cancer. This sequence represents the androgen receptor
CC transactivation signalling pathway modulator AB043547 described in the
CC method of the invention
XX
SQ Sequence 160820 BP; 46593 A; 31350 C; 32677 G; 50200 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 8; Length 160820;
Best Local Similarity 87.0%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 55080 TGAGAAGGCGAGTAGGAAGCTTTG 55102
RESULT 37
ADCS9504
ID ADCS9504 standard; cDNA; 1423 BP.
XX
AC ADCS9504;
XX
DT 18-DEC-2003 (first entry)
XX
DE Polypeptide-respiratory chain dehydrogenase-11.55 cDNA.
XX
KW Polypeptide-respiratory chain dehydrogenase-11.55; gene; ss; cancer;
KW malignant tumour; haemopathy; human immunodeficiency virus infection;
KW HIV; immunological disease; inflammation; antagonist.
XX
OS Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 225..542
XX /tag= a
XX /product= "Polypeptide-respiratory chain dehydrogenase-
XX 11.55"
XX /note= "This CDS is specifically claimed in claim 6"
XX
XX CNI1381565-A.
XX
XX 27-NOV-2002.
XX
XX 18-APR-2001; 2001CN-00112624.
XX
XX 18-APR-2001; 2001CN-00112624.
PR
```



PT developing a diagnostic marker or medicines for regulating their  
PT expression and activity, or as a target of gene therapy.

PS Claim 1; SEQ ID NO 1480; 305pp; English.

XX The invention relates to a novel human polynucleotide and the encoded  
XX polypeptide. A polynucleotide of the invention may have a use in gene  
XX therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
XX as a primer for synthesizing the polynucleotide or as a probe for  
XX detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are  
XX useful in gene therapy, for developing a diagnostic marker or medicines  
XX for regulating their expression and activity, or as a target of gene  
XX therapy. The proteins ADM01759-ADM06201 encoded by the polynucleotides  
XX are useful as pharmaceutical agents. The present sequence represents a  
XX cDNA sequence of the invention.

SQ Sequence 2533 BP; 590 A; 614 C; 707 G; 622 T; 0 U; 0 Other;

Query Match 77.4%; Score 17.8; DB 11; Length 2533;  
Best Local Similarity 90.5%; Pred. No. 4.8e+02; Length 2533;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 702 TGAGAAGGCAGTAGAAGCTT 682

RESULT 40  
ABL13110  
ID ABL13110 standard; cDNA; 6805 BP.

XX ABL13110;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33812.

XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; sb.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB69007.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signaling and cell-cell

XX interactions.

XX Claim 1; SEQ ID NO 33812; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 6805 BP; 2201 A; 1410 C; 1307 G; 1887 T; 0 U; 0 Other;

SQ Query Match 77.4%; Score 17.8; DB 4; Length 6805;

Best Local Similarity 90.5%; Pred. No. 5.3e+02; Length 6805;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTT 21

|||||  
Db 2650 TGAGAAGGCAGTAGAAGCTT 2670

RESULT 41

ABAL5166

ID ABA15166 standard; DNA; 9722 BP.

XX ABA15166;

XX 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 7497.

XX Human; neuroprotective; cytostatic; dermatological; virucide;  
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
XX antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;  
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US001334.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

XX 18-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226686P.

XX 22-AUG-2000; 2000US-0227182P.

PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0228287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229511P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250391P.
PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0251160P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232400P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234223P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234997P.	XX		
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239335P.	XX		
PR	13-OCT-2000;	2000US-0239337P.	XX		
PR	20-OCT-2000;	2000US-0240960P.	XX		
PR	20-OCT-2000;	2000US-0241785P.	XX		
PR	20-OCT-2000;	2000US-0241786P.	XX		
PR	20-OCT-2000;	2000US-0241787P.	XX		
PR	20-OCT-2000;	2000US-0241808P.	XX		
PR	20-OCT-2000;	2000US-0241809P.	XX		
PR	20-OCT-2000;	2000US-0241826P.	XX		
PR	01-NOV-2000;	2000US-0242221P.	XX		
PR	08-NOV-2000;	2000US-0246474P.	XX		
PR	08-NOV-2000;	2000US-0246475P.	XX		
PR	08-NOV-2000;	2000US-0246476P.	XX		
PR	08-NOV-2000;	2000US-0246477P.	XX		
PR	08-NOV-2000;	2000US-0246478P.	XX		
PR	08-NOV-2000;	2000US-0246523P.	XX		
PR	08-NOV-2000;	2000US-0246524P.	XX		
PR	08-NOV-2000;	2000US-0246525P.	XX		
PR	08-NOV-2000;	2000US-0246526P.	XX		
PR	08-NOV-2000;	2000US-0246527P.	XX		
PR	08-NOV-2000;	2000US-0246528P.	XX		
PR	08-NOV-2000;	2000US-0246532P.	XX		
PR	08-NOV-2000;	2000US-0246533P.	XX		
PR	08-NOV-2000;	2000US-0246610P.	XX		
PR	08-NOV-2000;	2000US-0246611P.	XX		
PR	08-NOV-2000;	2000US-0246613P.	XX		
PR	17-NOV-2000;	2000US-0249207P.	XX		
PR	17-NOV-2000;	2000US-0249208P.	XX		
PR	17-NOV-2000;	2000US-0249209P.	XX		

Query Match 77.4%; Score 17.8; DB 5; Length 9722;

Best Local Similarity 90.5%; Pred. No. 5.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTT 21  
|||||

Db 3027 TGAGAGGCAGTCGGAGCTT 3047  
|||||

RESULT 42  
ABA15164

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.

Disclosure; SEQ ID NO 7497; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (AB114678-AB118001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 9722 BP; 2169 A; 2302 C; 2639 G; 2612 T; 0 U; 0 Other;

ID ABAL15164 standard; DNA; 14855 BP.  
XX ABAL15164;  
AC  
XX  
XX  
DT 23-JAN-2002 (first entry)  
XX  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 7495.  
XX  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisclerotic; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
PN  
XX  
XX 16-AUG-2001.  
PD  
XX  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198121P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231988P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241836P.  
PR 20-OCT-2000; 2000US-0242211P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.



```
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 7495; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB11004-ABA21534) and proteins
CC (AB114678-AB118001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, diabetes mellitus, Crohn's disease,
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 14855 BP; 3335 A; 3554 C; 4013 G; 3953 T; 0 U; 0 Other;
SQ
Query Match 77.4%; Score 17.8; DB 5; Length 14855;
Best Local Similarity 90.5%; Pred. No. 5.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTT 21
Db 8160 TGAGAAGGCAGTGGGAAGCTT 8180

RESULT 43
ABA93286
ID ABA93286 standard; DNA; 22028 BP.
XX
XX ABA93286;
AC
XX
XX 22-APR-2002 (first entry)
DT
XX
XX Human acetyl-Coenzyme A acyltransferase 1 gene SEQ ID NO:1.
DE
XX
XX Human: acetyl-Coenzyme A acyltransferase; ACAA1; chromosome 3p23-p22;
KW peroxisomal 3-oxoacyl-Coenzyme A thiolase; SNP; genotype; haplotype;
KW single nucleotide polymorphism; polymorphic variant; enzyme; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH
```

```
FT variation
FT replace(3546,A)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(3537,A)
FT /*tag= b
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(4033,T)
FT /*tag= c
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(4157,T)
FT /*tag= d
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(4176,A)
FT /*tag= e
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(4276,C)
FT /*tag= f
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(9110,G)
FT /*tag= g
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(9182,A)
FT /*tag= h
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(9600,G)
FT /*tag= i
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(11702,T)
FT /*tag= j
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(11756,C)
FT /*tag= k
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(13390,T)
FT /*tag= l
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(14441,G)
FT /*tag= m
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(14872,C)
FT /*tag= n
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(14931,C)
FT /*tag= o
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(15471,C)
FT /*tag= p
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(15486,G)
FT /*tag= q
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(18004,T)
FT /*tag= r
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT replace(18192,T)
FT /*tag= s
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
XX WO200187903-A2.
XX
XX 22-NOV-2001.
XX
XX 03-MAY-2001; 2001WO-US014330.
XX
XX 18-MAY-2000; 2000US-0205022P.
XX (GENA-) GENAISSANCE PHARM INC.
XX (DUDA/) DUDA A E.
XX Chew A, Koshy B;
XX WPI; 2002-164134/21.
XX
```

PT Isolated polynucleotide, comprising a polymorphic variant of the acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase) gene useful for providing haplotype information and in therapy for treating related disorders.

XX Claim 19; Fig 1A-I; 93pp; English.

XX The present invention describes a polypeptide (I) which is a polymorphic variant (PV) of the acetyl-Coenzyme A acyltransferase (peroxisomal 3-oxoacyl-Coenzyme A thiolase) ACAAL protein (ABR05516). ACAAL is located on chromosome 3p23-p22. (I) can be encoded by ABA93286 (or ABA93288) CC where the sequence comprises one of the haplotypes shown in Table 4 or CC one of the haplotype pairs shown in Table 3, where Tables 3 and 4 are CC given in the specification. The polynucleotide encoding ACAAL can be used CC for providing haplotype and genotype information of an individual. CC Furthermore, the polynucleotide is useful for the treatment of disorders CC related to its abnormal expression or function. The present sequence CC represents the human ACAAL gene, featuring all the single nucleotide CC polymorphisms (SNP) possible, from the present invention

XX SQ Sequence 22028 BP; 4811 A; 5599 C; 6137 G; 5481 T; 0 U; 0 Other;

Query Match 77.4%; Score 17.8; DB 6; Length 22028;  
Best Local Similarity 90.5%; Pred. No. 5.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 12094 TGAGAAGGCAGTGGGAGCTT 12114

RESULT 44  
ABA93288  
ID ABA93288 standard; DNA; 22028 BP.

XX AC ABA93288;

XX DT 22-APR-2002 (first entry)

XX DE Human acetyl-Coenzyme A acyltransferase 1 gene SEQ ID NO:99.

XX KW Human; acetyl-Coenzyme A acyltransferase; ACAAL; chromosome 3p23-p22;  
KW peroxisomal 3-oxoacyl-Coenzyme A thiolase; SNP; genotype; haplotype;  
KW single nucleotide polymorphism; polymorphic variant; enzyme; gene; ds.

XX OS Homo sapiens.

XX PN WO200187903-A2.

XX PD 22-NOV-2001.

XX PF 03-MAY-2001; 2001WO-US014330.

XX PR 18-MAY-2000; 2000US-0205022P.

XX PA (GENA-) GENAISSANCE PHARM INC.  
PA (DUDA/) DUDA A E.

XX PI Chew A, Koshiy B;

XX DR WPI; 2002-164134/21.

XX Isolated polynucleotide, comprising a polymorphic variant of the acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase) gene useful for providing haplotype information and in therapy for treating related disorders.

XX PS Disclosure; Page 85-93; 93pp; English.

XX The present invention describes a polypeptide (I) which is a polymorphic variant (PV) of the acetyl-Coenzyme A acyltransferase (peroxisomal 3-oxoacyl-Coenzyme A thiolase) ACAAL protein (ABR05516). ACAAL is located CC on chromosome 3p23-p22. (I) can be encoded by ABA93286 (or ABA93288)

CC where the sequence comprises one of the haplotypes shown in Table 4 or CC one of the haplotype pairs shown in Table 3, where Tables 3 and 4 are CC given in the specification. The polynucleotide encoding ACAAL can be used CC for providing haplotype and genotype information of an individual. CC Furthermore, the polynucleotide is useful for the treatment of disorders CC related to its abnormal expression or function. The present sequence CC represents the human ACAAL gene, where all the single nucleotide CC polymorphisms (SNP) possible have been given in the sequence using the CC standard ambiguity codes

SQ Sequence 22028 BP; 4808 A; 5592 C; 6133 G; 5476 T; 0 U; 19 Other;

Query Match 77.4%; Score 17.8; DB 6; Length 22028;  
Best Local Similarity 90.5%; Pred. No. 5.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 12094 TGAGAAGGCAGTGGGAGCTT 12114

RESULT 45  
ABD33148/c

ID ABD33148 standard; DNA; 104399 BP.

XX AC ABD33148;

XX DT 18-NOV-2004 (first entry)

XX DE Murine cancer-associated (CA) gene MD07-019.

XX KW Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
KW ds; cancer; cytostatic.

XX OS Mus musculus.

XX PN WO2004058146-A2.

XX PD 15-JUL-2004.

XX PF 15-DEC-2003; 2003WO-US040081.

XX PR 17-DEC-2002; 2002US-00322281.

XX PA (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Malandro MS;

XX DR WPI; 2004-499109/47.

XX Novel human cancer associated protein encoded within open reading frame of cancer associated gene, useful as targets for diagnosing cancer.

XX PS Disclosure; SEQ ID NO 123; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-associated (CA) nucleic acids encoding them. The invention also relates to a method for treating cancers involving administering to a patient an inhibitor of CAP, and a method of screening for anticancer activity in a potential drug involving providing a cell that expresses a CA gene, contacting a tissue sample derived from a cancer cell with an anticancer drug candidate and monitoring the effect of the anticancer drug candidate on expression of the CA gene. The CAP proteins are useful for detecting cancer associated with expression of a CAP protein in a test cell sample and for screening for a bioactive agent capable of modulating the CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing CC cancer, involving determining the expression of a CA nucleic acid in a CC tissue. This sequence represents a murine CA gene of the invention. Note: CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 104399 BP; 25540 A; 20083 C; 20918 G; 29473 T; 0 U; 8385 Other;

```
Query Match      77.4%; Score 17.8; DB 13; Length 104399;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GAGAAAGGCAGTAGAAAGCTTAG 22
DB      35008 GAGAAAGGCAGTAGAAAGACTA 34988

RESULT 46
AAV21209_12/c
Continuation (13 of 17) of AAV21209 from base 1200001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name      Begin      End
WP AAV21209_00        1      110000
WP AAV21209_01      100001      210000
WP AAV21209_02      200001      310000
WP AAV21209_03      300001      410000
WP AAV21209_04      400001      510000
WP AAV21209_05      500001      610000
WP AAV21209_06      600001      710000
WP AAV21209_07      700001      810000
WP AAV21209_08      800001      910000
WP AAV21209_09      900001     1010000
WP AAV21209_10     1000001     1110000
WP AAV21209_11     1100001     1210000
WP AAV21209_12     1200001     1310000
WP AAV21209_13     1300001     1410000
WP AAV21209_14     1400001     1510000
WP AAV21209_15     1500001     1610000
WP AAV21209_16     1600001     1664976

Query Match      77.4%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
DB      105980 AGAAGGCAGTAGAAAGCTTAG 105960

RESULT 47
AAV21209_13/c
Continuation (14 of 17) of AAV21209 from base 1300001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name      Begin      End
WP AAV21209_00        1      110000
WP AAV21209_01      100001      210000
WP AAV21209_02      200001      310000
WP AAV21209_03      300001      410000
WP AAV21209_04      400001      510000
WP AAV21209_05      500001      610000
WP AAV21209_06      600001      710000
WP AAV21209_07      700001      810000
WP AAV21209_08      800001      910000
WP AAV21209_09      900001     1010000
WP AAV21209_10     1000001     1110000
WP AAV21209_11     1100001     1210000
WP AAV21209_12     1200001     1310000
WP AAV21209_13     1300001     1410000
WP AAV21209_14     1400001     1510000
WP AAV21209_15     1500001     1610000
WP AAV21209_16     1600001     1664976

Query Match      77.4%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
DB      5980 AGAAGGCAGTAGAAAGCTTAG 5960
```

```
RESULT 48
ADD68802
ID ADD68802 standard; DNA; 21 BP.
XX
XX AC ADD68802;
XX
XX DT 15-JAN-2004 (first entry)
XX
XX DE PCR primer ptsU used to amplify Group A Streptococcus ptsI sequences.
XX
XX KW beta-haemolytic Group A Streptococcus; GAS;
XX KW fluorescence resonance energy transfer; FRET; ptsU; PCR; primer; ptsI;
XX KW ss.
XX
XX OS Streptococcus sp. 'group A'.
XX
XX PN US6593093-B1.
XX
XX PD 15-JUL-2003.
XX
XX PF 20-FEB-2002; 2002US-00081923.
XX
XX PR 20-FEB-2002; 2002US-00081923.
XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX PI Uhl JR, Cockerill FR;
XX
XX DR WPI; 2003-828265/77.
XX
XX PT Detecting presence/absence of group A Streptococcus in sample by
XX PT performing cycling step comprising amplifying, hybridizing DNA from
XX PT sample by fluorescent probe, detecting for fluorescent resonance energy
XX PT transfer.
XX
XX PS Claim 2; SEQ ID NO 1; 35pp; English.
XX
XX CC The invention relates to a novel method for detecting the presence or
XX CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX CC isolated from an individual by performing more than one cycling step. The
XX CC method involves an amplifying step comprising contacting the sample with
XX CC GAS primers, a subsequent hybridizing step comprising contacting the
XX CC amplified product with fluorescent-labelled probes and finally, detecting
XX CC fluorescence resonance energy transfer (FRET) where the presence or
XX CC absence of FRET indicates the presence or absence of GAS in the sample.
XX CC The method of the invention may be useful for detecting the presence or
XX CC absence of GAS in a biological sample from an individual. The current
XX CC sequence is that of the ptsU PCR primer of the invention which was used
XX CC to amplify GAS ptsI sequences.
XX
XX SQ Sequence 21 BP; 9 A; 2 C; 6 G; 4 T; 0 U; 0 Other;

Query Match      75.7%; Score 17.4; DB 10; Length 21;
Best Local Similarity 94.7%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 AAGGCAGTAGAAAGCTTAG 23
DB      2 AATGCAGTAGAAAGCTTAG 20

RESULT 49
ADJ61989
ID ADJ61989 standard; DNA; 21 BP.
XX
XX AC ADJ61989;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE ptsI gene PCR primer, ptsU.
XX
XX KW Group A Streptococcus; GAS; PCR; primer; ss.
```



CC compounds or the progression of these toxic effects by determining the  
CC global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic effect  
CC of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression. The  
CC method can also be used to identify an agent which modulates the toxic  
CC response and predict cellular pathways that a compound modulates in a  
CC cell. The methods utilize a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridises to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cells exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
CC characterised by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent  
XX  
SQ Sequence 322 BP; 96 A; 73 C; 74 G; 79 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 6; Length 322;  
Best Local Similarity 94.7%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAAAGC 19  
|||  
DB 252 TGAGAGGCGAGTAAAGC 234  
|||

RESULT 52  
ADZ62746  
ID ADZ62746 standard; DNA; 2151 BP.  
XX  
AC ADZ62746;  
XX  
DT 30-JUN-2005 (first entry)  
XX  
DE Murine Cdc45i gene, SEQ ID 2470.  
XX  
KW Drug screening; Peroxisome Proliferator-Activated Receptor alpha;  
XX PPAR alpha; gene; ds.  
XX  
OS Mus musculus.  
XX  
PN US2005084872-A1.  
XX  
PD 21-APR-2005.  
XX  
PF 23-JAN-2004; 2004US-00764420.  
XX  
PR 24-JAN-2003; 2003US-0442797P.  
PR 30-MAY-2003; 2003US-0474413P.  
XX  
PA (LUMP/) LUM P Y.  
PA (TANY/) TAN Y.  
PA (DAIH/) DAI H.  
PA (MUIS/) MUISE E S.  
PA (BERG/) BERGER J P.  
PA (THOM/) THOMPSON J R.  
XX  
PI Lum PY, Tan Y, Dai H, Muise ES, Berger JP, Thompson JR;

XX WPI; 2005-313963/32.  
DR REFSEQ; NM\_009862.  
XX  
PT Determining whether agent has defined biological activity, by making  
PT comparison e.g. comparing efficacy value, toxicity value and classifier  
PT value of agent to reference values, and using results to determine  
PT whether agent has activity.  
XX  
PS Example 4; SEQ ID NO 2470; 51pp; English.  
XX  
CC The present invention relates to a methods for screening biologically  
CC active agents, such as candidate drugs, to determine whether the agent  
CC possesses a defined biological activity. The methods involve making one  
CC or more comparisons chosen from comparing efficacy value of agent to  
CC reference efficacy value, comparing a toxicity value of the agent to  
CC reference toxicity value, comparing a classifier value of the agent to  
CC reference classifier value, and using the comparison result(s) to  
CC determine whether the agent possesses the defined biological activity.  
CC Also claimed is a population of oligonucleotide probes (1) specific for  
CC measuring the expression levels of members of a classifier population of  
CC genes or a toxicity-related population of genes. (1) are useful for  
CC measuring the expression levels of genes that are useful for identifying  
CC agonists or partial agonists of Peroxisome Proliferator-Activated  
CC Receptor (PPAR) gamma. ADZ63072-ADZ63959 are oligonucleotide probes which  
CC are useful in the method of the invention to measure the expression  
CC pattern of mouse genes ADZ62273-ADZ63071 that yield an expression pattern  
CC that correlates with the stimulation of PPARalpha receptors by an agent.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=20050084872.  
XX  
SQ Sequence 2151 BP; 559 A; 499 C; 551 G; 542 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 14; Length 2151;  
Best Local Similarity 94.7%; Pred. No. 7.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAGAAGCGCAGTAGAAAGC 19  
|||  
DB 1891 TCAGAAGCGCAGTAGAAAGC 1909  
|||

RESULT 53  
AAK65192/C  
ID AAK65192 standard; DNA; 7613 BP.  
XX  
AC AAK65192;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20004.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-02114886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0211647P.  
PR 07-JUL-2000; 2000US-02116880P.  
PR 11-JUL-2000; 2000US-02117487P.  
PR 11-JUL-2000; 2000US-02117496P.  
PR 14-JUL-2000; 2000US-02118290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0228279P.  
PR 22-AUG-2000; 2000US-0228681P.  
PR 22-AUG-2000; 2000US-0228688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX PS Disclosure; SEQ ID NO 20004; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 XX  
 SQ Sequence 7613 BP; 1434 A; 2458 C; 2149 G; 1572 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 4; Length 7613;  
 Best Local Similarity 94.7%; Pred. No. 8.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTT 21  
 |||||  
 Db 3366 AGAAGGCAGAGAAAGCTT 3348

RESULT 54  
 ABL28640  
 ID ABL28640 standard; DNA; 24789 BP.  
 XX  
 AC ABL28640;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37393.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Claim 1; SEQ ID NO 37393; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 24789 BP; 7654 A; 5448 C; 6281 G; 5406 T; 0 U; 0 Other;  
 SQ  
 Query Match 75.7%; Score 17.4; DB 4; Length 24789;  
 Best Local Similarity 94.7%; Pred. No. 9e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TGAGAAGGCAGTAGAAGC 19  
 |||||  
 Db 13733 TGAGAAGGCAGAGAAAGC 13751

RESULT 55  
 ABK62529  
 ID ABK62529 standard; cDNA; 519 BP.

XX  
 AC ABK62529;  
 XX

XX 18-JUN-2002 (first entry)

XX Rat sequence differentially expressed in response to a hepatotoxin #436.  
 DE Rat sequence differentially expressed in response to a hepatotoxin #436.  
 XX

KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
 KW differential expression; centrilobular necrosis; steatosis.

OS Rattus norvegicus.

XX  
 PN WO200210453-A2.

XX 07-FEB-2002.

XX 30-JUL-2001; 2001WO-US023872.

XX 31-JUL-2000; 2000US-0222040P.

PR 02-NOV-2000; 2000US-0244880P.

PR 11-MAY-2001; 2001US-0290029P.

PR 15-MAY-2001; 2001US-0290645P.

PR 22-MAY-2001; 2001US-0292336P.

PR 06-JUN-2001; 2001US-0295798P.

PR 13-JUN-2001; 2001US-0297457P.

PR 19-JUN-2001; 2001US-0298884P.

PR 09-JUL-2001; 2001US-0303459P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

PI WPI; 2002-241625/29.

XX Predicting toxic effects of compounds or the progression of these toxic  
 PT effects by determining the changes in gene expression in tissues or cells  
 PT exposed to the toxin and comparing these to gene expression in unexposed  
 PT tissues or cells.

XX Claim 1; SEQ ID NO 436; 239pp; English.  
 CC The invention relates to methods for predicting toxic effects of  
 CC compounds or the progression of these toxic effects by determining the  
 CC global changes in gene expression in tissues or cells exposed to the  
 CC toxin and comparing these to gene expression in unexposed tissues or  
 CC cells. Also included are methods of predicting at least one toxic effect  
 CC of a compound or progression of a toxic effect, preferably the  
 CC hepatotoxicity of a compound, comprising detecting the level of  
 CC expression in a tissue or cell sample exposed to the compound of two or  
 CC more genes listed in the specification, where differential expression of  
 CC the genes is indicative of at least one toxic effect or progression. The  
 CC method can also be used to identify an agent which modulates the toxic  
 CC response and predict cellular pathways that a compound modulates in a  
 CC cell. The methods utilise a set of at least two probes (on a solid  
 CC support in kit form), where each of the probes comprises a sequence that  
 CC specifically hybridises to a gene listed in the specification, a computer  
 CC system comprising a database containing information identifying the  
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
 CC set of genes comprising at least two genes listed in the specification,  
 CC and a user interface to view the information used to present information,  
 CC identifying the expression level in a tissue or cell of at least one gene  
 CC listed in the specification. The method is useful for elucidating global  
 CC changes in gene expression and for identifying toxicity markers in

CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
CC characterized by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent  
XX  
XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 56  
ADBS6078  
ID ADBS6078 standard; DNA; 519 BP.

AC ADBS6078;

XX 04-DEC-2003 (first entry)

DE Toxicity-related gene, SEQ ID 1104.

XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
KW drug screening; toxicity assay; ds.

XX Unidentified.

XX WO2003064624-A2.

XX 07-AUG-2003.

XX 31-JAN-2003; 2003WO-US003194.

XX 31-JAN-2002; 2002US-00060087.

PR 15-MAR-2002; 2002US-0364045P.

PR 15-MAR-2002; 2002US-0364055P.

PR 30-DEC-2002; 2002US-0436643P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI; 2003-689530/65.

XX Predicting a toxic effect of a compound, useful in identifying toxicity  
XX markers in liver tissues or cells for drug screening and toxicity assays,  
XX comprises preparing gene expression profile of tissue or cells exposed to  
XX the compound.

XX Claim 1; SEQ ID NO 1104; 1156pp; English.

XX The present invention relates to a method for predicting a toxic effect  
XX of a compound. The method comprises preparing a gene expression profile  
XX of a tissue or cell sample exposed to the compound, and comparing the  
XX gene expression profile to a database comprising SEQ ID 1-4925, where  
XX differential expression of the gene indicates at least one toxic effect.  
XX The method is useful for predicting at least one toxic effect of a  
XX compound, predicting hepatotoxicity or the progression of a toxic effect  
XX of a compound, identifying an agent that modulates the onset or  
XX progression of a toxic response, predicting the cellular pathways that a  
XX compound modulates in a cell, and identifying an agent that modulates at  
XX least one activity of a protein. The method and compositions of the  
XX present invention using a database of genes having liver toxin-induced  
XX differential expression, are useful in identifying toxicity markers in  
XX liver tissues or cells for drug screening and toxicity assays. Note: The

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 11-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 11-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 11-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 11-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 11-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGCAGTAAAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.



CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.  
XX SQ Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e-02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GAGAAGGCAGTAGAAGCTTAG 23  
DB 179 GGGAAAGCAGTAGAAGCTTAG 200

RESULT 58  
ABL90230/c  
ID ABL90230 standard; cDNA; 670 BP.

XX AC ABL90230;

XX DT 24-MAY-2002 (first entry)

XX DE Human polynucleotide SEQ ID NO 792.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.

XX OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US016450.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX DR WPI; 2002-122018/16.

XX DR P-PSDB; ABB98821.

XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders.

XX PS Claim 4; SEQ ID NO 792; 2081pp + Sequence Listing; English.

XX CC The invention relates to novel genes (ABL9449-ABL90853) and proteins  
CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 670 BP; 182 A; 133 C; 115 G; 229 T; 0 U; 11 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 670;  
Best Local Similarity 86.4%; Pred. No. 7.9e-02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GAGAAGGCAGTAGAAGCTTAG 23  
DB 549 GAGAAGGCAGTAGAAGGAGAG 528

RESULT 59  
AAX97673/c  
ID AAX97673 standard; DNA; 886 BP.

XX AC AAX97673;

XX DT 13-SEP-1999 (first entry)

XX DE Extended human secreted protein coding sequence, SEQ ID NO. 137.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
KW cellular differentiation; immune system regulator; anti-inflammatory;  
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
KW genetic disease; ss.

XX OS Homo sapiens.

XX PN WO9311236-A2.

XX PD 24-JUN-1999.

XX PF 17-DEC-1998; 98WO-IB002122.

XX PR 17-DEC-1997; 97US-0069957P.

XX PR 09-FEB-1998; 98US-0074121P.

XX PR 13-APR-1998; 98US-0081563P.

XX PR 10-AUG-1998; 98US-0096116P.

XX PA (GEST ) GENSET.

XX PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX DR WPI; 1999-385906/32.

XX DR P-PSDB; AAY35989.

XX PT New isolated human secreted proteins.

XX PS Claim 1; Page 265-266; 516pp; English.

XX CC This sequence represents an extended human secreted protein coding  
CC sequence of the invention. The secreted proteins can be used in treating  
CC or controlling a variety of human conditions. The secreted proteins may  
CC act as cytokines or may affect cellular proliferation or differentiation  
CC or may act as immune system regulators, haematopoiesis regulators, tissue  
CC growth regulators, regulators of reproductive hormones or cell movement  
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or  
CC tumour inhibition activity. The DNAs can be used in forensic procedures  
CC to identify individuals or in diagnostic procedures to identify  
CC individuals having genetic diseases resulting from abnormal expression of  
CC the genes corresponding to the extended cDNAs. They are also useful for  
CC constructing a high resolution map of the human chromosomes. They can  
CC also be used for gene therapy to control or treat genetic diseases



GenCore version 5.1.7  
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 27, 2006, 20:43:21 ; Search time 196,315 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-2  
Perfect score: 21  
Sequence: 1 tgcattgattgggtattcttcc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues  
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : N Geneseq\_21.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	ADD68803	ADD68803 PCR prime
2	21	100.0	21	ADJ61990	Adj61990 PtsI gene
3	21	100.0	21	ADZ75862	Adz75862 Group B s
C 4	21	100.0	1323	10 ADD68796	ADD68796 Streptoco
C 5	21	100.0	1323	12 ADJ62000	Adj62000 Group A s
C 6	21	100.0	1378	10 ADD68798	ADD68798 Streptoco
C 7	21	100.0	1378	12 ADJ62002	Adj62002 Group A s
C 8	21	100.0	1379	10 ADD68800	ADD68800 Streptoco
C 9	21	100.0	1379	10 ADD68797	ADD68797 Streptoco
C 10	21	100.0	1379	10 ADD68791	ADD68791 Streptoco
C 11	21	100.0	1379	12 ADJ62001	Adj62001 Group A s
C 12	21	100.0	1379	12 ADJ62004	Adj62004 Group A s
C 13	21	100.0	1379	12 ADJ61995	Adj61995 Group A s
C 14	21	100.0	1384	10 ADD68792	ADD68792 Streptoco
C 15	21	100.0	1384	10 ADD68794	ADD68794 Streptoco
C 16	21	100.0	1384	12 ADJ61998	Adj61998 Group A s
C 17	21	100.0	1384	12 ADJ61996	Adj61996 Group A s
C 18	21	100.0	1385	10 ADD68793	ADD68793 Streptoco
C 19	21	100.0	1385	12 ADJ61997	Adj61997 Group A s

C 20	21	100.0	1390	10 ADD68795	ADD68795 Streptoco
C 21	21	100.0	1390	12 ADJ61999	Adj61999 Group A s
C 22	21	100.0	1393	10 ADD68799	ADD68799 Streptoco
C 23	21	100.0	1393	12 ADJ62003	Adj62003 Group A s
C 24	21	100.0	1447	10 ADD68801	ADD68801 Streptoco
C 25	21	100.0	1447	12 ADJ62005	Adj62005 Group A s
C 26	21	100.0	1731	6 ABN68277	ABN68277 Streptoco
C 27	21	100.0	1731	6 ABN68276	ABN68276 Streptoco
C 28	21	100.0	1734	8 ACA50572	ACA50572 Prokaryot
C 29	21	100.0	1734	13 ADV85417	ADV85417 Streptoco
C 30	21	100.0	1803	10 ADD68806	ADD68806 Streptoco
C 31	21	100.0	1803	12 ADJ61993	Adj61993 Group A s
C 32	21	100.0	29072	13 ADV87723	ADV87723 Streptoco
C 33	21	100.0	29072	13 ADV78976	ADV78976 Streptoco
C 34	21	100.0	110000	6 ABN71527_07	Continuation (8 of
C 35	21	100.0	110000	13 ADV81204_08	Continuation (9 of
C 36	19.4	92.4	402	10 ABX06888	ABX06888 S. pneumo
C 37	19.4	92.4	962	2 AAV65250	AAV65250 DNA encod
C 38	19.4	92.4	1140	13 ADK44676	ADK44676 Streptoco
C 39	19.4	92.4	1287	13 ADR92834	ADR92834 Novel S.
C 40	19.4	92.4	1287	14 AEA56704	AEA56704 Streptoco
C 41	19.4	92.4	1734	4 AAS55546	AAS55546 Streptoco
C 42	19.4	92.4	1734	8 ACA49934	ACA49934 Prokaryot
C 43	19.4	92.4	8195	2 AAV52227	AAV52227 Streptoco
C 44	19.4	92.4	110000	10 ABS56454_09	Continuation (10 o
C 45	19.4	92.4	110000	10 ABS56454_10	Continuation (11 o
C 46	18.4	87.6	553	12 ACH68181	ACH68181 Human gen
C 47	17.8	84.8	1830	9 ADB11789	ADB11789 Alloiococ
C 48	17.8	84.8	1830	9 ADB11785	ADB11785 Alloiococ
C 49	17.8	84.8	1830	9 ADB11787	ADB11787 Alloiococ
C 50	17.8	84.8	1830	9 ADR11783	ADR11783 Alloiococ
C 51	17.8	84.8	2914	7 ADR32245	ADR32245 Human tum
C 52	17.8	84.8	2914	10 ADA53346	ADA53346 Human cod
C 53	17.8	84.8	110000	9 ADB12064_13	Continuation (14 o
C 54	16.8	80.0	412	6 ABK74374	ABK74374 Bacillus
C 55	16.8	80.0	1464	2 AAX30475	Aax30475 H. pylori
C 56	16.8	80.0	1554	2 AAT68140	Aat68140 H. pylori
C 57	16.8	80.0	2547	12 ADE91329	Ade91329 Maize glu
C 58	16.8	80.0	2953	6 ABV72307	ABV72307 Nucleotid
C 59	16.8	80.0	48179	14 ADZ12700	Adz12700 Murine ca
C 60	16.8	80.0	51261	9 ADA02873	Ada02873 Mouse Dpt

ALIGNMENTS

RESULT 1  
ADD68803

ID ADD68803 standard; DNA; 21 BP.

XX AC ADD68803;

XX DT 15-JAN-2004 (first entry)

XX PCR primer ptsI used to amplify Group A Streptococcus ptsI sequences.

DE beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptsI; PCR; primer; ptsI;  
KW ss.

XX Streptococcus sp. 'group A'.

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

PI

XX WPI; 2003-828265/77.  
XX Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
XX Claim 2; SEQ ID NO 2; 35pp; English.  
XX  
XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the ptstL PCR primer of the invention which was used  
CC to amplify GAS ptstL sequences.  
XX  
XX Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 21; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 1 TGCATGTATGGGTATCTTCC 21  
|||||

RESULT 2  
ADJ61990  
ID ADJ61990 standard; DNA; 21 BP.  
XX  
XX AC ADJ61990;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX PtsI gene PCR primer, ptstL.  
XX  
XX Group A Streptococcus; GAS; PCR; primer; ss.  
XX  
XX Streptococcus sp.  
XX  
XX US2004014118-A1.  
XX  
XX 22-JAN-2004.  
XX  
XX 19-JUN-2003; 2003US-00465205.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
XX Uhl JR, Cockerill FR;  
XX  
XX WPI; 2004-224229/21.  
XX  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptstL primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
XX Claim 2; SEQ ID NO 2; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptstL primers and probes to produce ptstL  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of

CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a ptstL gene PCR primer.  
XX  
XX Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 21; DB 12; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 1 TGCATGTATGGGTATCTTCC 21  
|||||

RESULT 3  
ADZ75862  
ID ADZ75862 standard; DNA; 21 BP.  
XX  
XX AC ADZ75862;  
XX  
XX 28-JUL-2005 (first entry)  
XX  
XX Group B streptococcus phosphotransferase (pts) PCR primer, SEQ ID NO:2.  
XX  
XX Microorganism detection; fluorescence; diagnosis;  
KW streptococcus infection; infection; gynecology and obetetrice;  
KW phosphotransferase; PCR; primer; ss.  
XX  
XX Streptococcus sp. 'group B'.  
XX  
XX US2005106578-A1.  
XX  
XX 19-MAY-2005.  
XX  
XX 18-NOV-2003; 2003US-00716005.  
XX  
XX 18-NOV-2003; 2003US-00716005.  
XX  
XX {UHLJ/} UHL J R.  
XX  
XX {COCK/} COCKERILL F R.  
XX  
XX {AICH/} AICHINGER C.  
XX  
XX {REIS/} REISER A.  
XX  
XX Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
XX  
XX WPI; 2005-371550/38.  
XX  
XX Detecting group B streptococcus, comprises amplifying a sample with pts  
PT primers, hybridizing the sample with fluorescently labeled pts probes,  
PT and detecting the presence of fluorescence resonance energy transfer.  
XX  
XX Claim 2; SEQ ID NO 2; 13pp; English.  
XX  
XX The invention relates to a real-time PCR-based method of detecting the  
CC presence or absence of group B streptococcus (GBS) bacterial pathogens in  
CC a biological sample from an individual. The method comprises  
CC amplification of a conserved region of the phosphotransferase (pts) gene  
CC of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the  
CC amplification product with a pair of fluorescently labeled  
CC phosphotransferase probes (especially ADZ75863-ADZ75864), and detection  
CC of the presence or absence of fluorescence resonance energy transfer  
CC (FRET), where presence of FRET is indicative of the presence of GBS in  
CC the sample. GBS infection is a leading cause of neonatal morbidity and  
CC mortality, with infection occurring during childbirth. Currently, it is  
CC recommended that women are screened for GBS during week 35-37 of  
CC gestation by a culture-based method which may take up to 72 hours for a  
CC result. However, many women first present at healthcare facilities at the  
CC time of labor, and in addition, GBS infection can be transient, so that a  
CC woman free of GBS at the time of screening may not be free of GBS when  
CC she is due to give birth. The method of the invention provides a real-

CC time assay for the detection of group B streptococcus in a sample, is  
CC more sensitive and specific than prior art non-culture based methods, and  
CC can thus be implemented for routine diagnosis of the presence of group B  
CC streptococcus. Sequences ADZ75861-ADZ75862 represent GAS  
CC phosphotransferase PCR primers specifically claimed for use in the method  
CC of the invention.

XX Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 21; DB 14; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21

Db 1 TGCATGTATGGGTTATCTTCC 21

#### RESULT 4

ADD68796/C

ID ADD68796 standard; DNA; 1323 BP.

XX AC

XX ADD68796;

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptst DNA - SEQ ID 12.

DE beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptst; ds.  
XX

OS Streptococcus sp. 'group A'.

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

DR Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 12; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridizing step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptst DNA (SEQ ID 12)  
XX of the invention.

XX Sequence 1323 BP; 395 A; 278 C; 297 G; 353 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 21; DB 10; Length 1323;

Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 193 TGCATGTATGGGTTATCTTCC 173

#### RESULT 5

ADD62000/C

ID ADJ62000 standard; DNA; 1323 BP.

XX AC

XX ADJ62000;

XX 06-MAY-2004 (first entry)

XX Group A streptococcus ptst gene #7.

XX Group A Streptococcus; GAS; gene; ds.

XX Streptococcus sp.

XX US2004014118-A1.

XX 22-JAN-2004.

XX 19-JUN-2003; 2003US-00465205.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2004-224229/21.

XX Detecting Group A Streptococcus in biological sample, involves contacting  
XX sample with pair of ptst primers and probes and detecting fluorescence  
XX resonance energy transfer indicating Group A Streptococcus in sample.  
XX Disclosure; SEQ ID NO 12; 36pp; English.

XX The invention relates to a method of detecting the presence or absence of  
XX Group A Streptococcus (GAS) in biological sample from individual, by  
XX contacting sample with pair of ptst primers and probes to produce ptst  
XX amplification product, where first, second probes are labelled with donor  
XX and acceptor fluorescent moiety, and detecting presence or absence of  
XX fluorescence resonance energy transfer, which indicates the presence or  
XX absence of GAS in biological sample. The method is useful for detecting  
XX the presence or absence of Group A Streptococcus in a biological sample  
XX such as throat swabs, tissues and bodily fluids from an individual. The  
XX present sequence represents a group A streptococcus ptst gene.

XX Sequence 1323 BP; 395 A; 278 C; 297 G; 353 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 21; DB 12; Length 1323;

Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21

Db 193 TGCATGTATGGGTTATCTTCC 173

#### RESULT 6

ADD68798/C

ID ADD68798 standard; DNA; 1378 BP.

XX AC

XX ADD68798;

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptst DNA - SEQ ID 14.

DE beta-haemolytic Group A Streptococcus; GAS;

XX fluorescence resonance energy transfer; FRET; ptst; ds.

KW

XX OS Streptococcus sp. 'group A'.

XX PN US6593093-B1.

XX PD 15-JUL-2003.

XX PF 20-FEB-2002; 2002US-00081923.

XX PR 20-FEB-2002; 2002US-00081923.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Uhl JR, Cockerill FR;

XX DR WPI; 2003-828265/77.

XX PT Detecting presence/absence of group A Streptococcus in sample by

XX PT performing cycling step comprising amplifying, hybridizing DNA from

XX PT sample by fluorescent probe, detecting for fluorescent resonance energy

XX PT transfer.

XX PS Example 2; SEQ ID NO 14; 35pp; English.

XX CC The invention relates to a novel method for detecting the presence or

XX CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample

XX CC isolated from an individual by performing more than one cycling step. The

XX CC method involves an amplifying step comprising contacting the sample with

XX CC GAS primers, a subsequent hybridising step comprising contacting the

XX CC amplified product with fluorescent-labelled probes and finally, detecting

XX CC fluorescence resonance energy transfer (FRET) where the presence or

XX CC absence of FRET indicates the presence or absence of GAS in the sample.

XX CC The method of the invention may be useful for detecting the presence or

XX CC absence of GAS in a biological sample from an individual. The current

XX CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 14)

XX CC of the invention.

XX SQ Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1378;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

DB 209 TGCATGTATGGGTATCTTCC 189

RESULT 7

ADJ62002/c

ID ADJ62002 standard; DNA; 1378 BP.

XX AC ADJ62002;

XX DT 06-MAY-2004 (first entry)

XX DE Group A streptococcus PtaI gene #9.

XX KW Group A Streptococcus; GAS; gene; ds.

XX OS Streptococcus sp.

XX XX US2004014119-A1.

XX PD 22-JAN-2004.

XX PF 19-JUN-2003; 2003US-00465205.

XX PR 20-FEB-2002; 2002US-00081923.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX XX Uhl JR, Cockerill FR;

XX DR WPI; 2004-224229/21.

XX XX

XX PT Detecting Group A Streptococcus in biological sample, involves contacting

XX PT sample with pair of ptai primers and probes and detecting fluorescence

XX PT resonance energy transfer indicating Group A Streptococcus in sample.

XX PS Disclosure; SEQ ID NO 14; 36pp; English.

XX XX

XX CC The invention relates to a method of detecting the presence or absence of

XX CC Group A Streptococcus (GAS) in biological sample from individual, by

XX CC contacting sample with pair of ptai primers and probes to produce ptai

XX CC amplification product, where first, second probes are labelled with donor

XX CC and acceptor fluorescent moiety, and detecting presence or absence of

XX CC fluorescence resonance energy transfer, which indicates the presence or

XX CC absence of GAS in biological sample. The method is useful for detecting

XX CC the presence or absence of Group A Streptococcus in a biological sample

XX CC such as throat swabs, tissues and bodily fluids from an individual. The

XX CC present sequence represents a group A streptococcus ptai gene.

XX SQ Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1378;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

DB 209 TGCATGTATGGGTATCTTCC 189

RESULT 8

ADD6800/c

ID ADD6800 standard; DNA; 1379 BP.

XX AC ADD6800;

XX DT 15-JAN-2004 (first entry)

XX XX

XX DE Streptococcus sp. 'group A' ptai DNA - SEQ ID 16.

XX KW beta-haemolytic Group A Streptococcus; GAS;

XX KW fluorescence resonance energy transfer; FRET; ptai; ds.

XX OS Streptococcus sp. 'group A'.

XX PN US6593093-B1.

XX PD 15-JUL-2003.

XX PF 20-FEB-2002; 2002US-00081923.

XX PR 20-FEB-2002; 2002US-00081923.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Uhl JR, Cockerill FR;

XX DR WPI; 2003-828265/77.

XX PT Detecting presence/absence of group A Streptococcus in sample by

XX PT performing cycling step comprising amplifying, hybridizing DNA from

XX PT sample by fluorescent probe, detecting for fluorescent resonance energy

XX PT transfer.

XX PS Example 2; SEQ ID NO 16; 35pp; English.

XX XX

XX CC The invention relates to a novel method for detecting the presence or

XX CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample

XX CC isolated from an individual by performing more than one cycling step. The

XX CC method involves an amplifying step comprising contacting the sample with

XX CC GAS primers, a subsequent hybridising step comprising contacting the

XX CC amplified product with fluorescent-labelled probes and finally, detecting

XX CC fluorescence resonance energy transfer (FRET) where the presence or

XX CC absence of FRET indicates the presence or absence of GAS in the sample.

XX CC The method of the invention may be useful for detecting the presence or

XX CC absence of GAS in a biological sample from an individual. The current

XX CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 14)

XX CC of the invention.

XX SQ Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;

CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 16)  
CC of the invention.  
XX  
SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 210 TGCATGTATGGTTATCTTCC 190  
  
RESULT 9  
ADD68797/c  
ID ADD68797 standard; DNA; 1379 BP.  
XX AC ADD68797;  
XX DT 15-JAN-2004 (first entry)  
XX DE Streptococcus sp. 'group A' ptaI DNA - SEQ ID 13.  
XX KW beta-haemolytic Group A Streptococcus; GAS;  
XX KW fluorescence resonance energy transfer; FRET; ptaI; ds.  
XX OS Streptococcus sp. 'group A'.  
XX PN US6593093-B1.  
XX PD 15-JUL-2003.  
XX PF 20-FEB-2002; 2002US-00081923.  
XX PR 20-FEB-2002; 2002US-00081923.  
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX PI Uhl JR, Cockerill FR;  
XX DR WPI; 2003-828265/77.  
XX PT Detecting presence/absence of group A Streptococcus in sample by  
XX performing cycling step comprising amplifying, hybridizing DNA from  
XX sample by fluorescent probe, detecting for fluorescent resonance energy  
XX transfer.  
XX PS Example 2; SEQ ID NO 13; 35pp; English.  
XX CC The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridising step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 13)  
XX of the invention.  
XX SQ Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189  
  
RESULT 10  
ADD68791/c  
ID ADD68791 standard; DNA; 1379 BP.  
XX AC ADD68791;  
XX DT 15-JAN-2004 (first entry)  
XX DE Streptococcus sp. 'group A' ptaI DNA - SEQ ID 7.  
XX KW beta-haemolytic Group A Streptococcus; GAS;  
XX KW fluorescence resonance energy transfer; FRET; ptaI; ds.  
XX OS Streptococcus sp. 'group A'.  
XX PN US6593093-B1.  
XX PD 15-JUL-2003.  
XX PF 20-FEB-2002; 2002US-00081923.  
XX PR 20-FEB-2002; 2002US-00081923.  
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX PI Uhl JR, Cockerill FR;  
XX DR WPI; 2003-828265/77.  
XX PT Detecting presence/absence of group A Streptococcus in sample by  
XX performing cycling step comprising amplifying, hybridizing DNA from  
XX sample by fluorescent probe, detecting for fluorescent resonance energy  
XX transfer.  
XX PS Example 2; SEQ ID NO 7; 35pp; English.  
XX CC The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridising step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 7)  
XX of the invention.  
XX SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189  
  
RESULT 11  
ADJ62001/c  
ID ADJ62001 standard; DNA; 1379 BP.  
XX AC ADJ62001;  
XX DT 06-MAY-2004 (first entry)  
XX

DE Group A streptococcus PtsI gene #8.  
 XX  
 KW Group A Streptococcus; GAS; gene; ds.  
 XX  
 OS Streptococcus sp.  
 XX  
 PN US2004014118-A1.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 XX  
 XX 20-FEB-2002; 2002US-00081923.  
 XX  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;  
 XX  
 PD WPI; 2004-224229/21.  
 XX  
 PT Detecting Group A Streptococcus in biological sample, involves contacting  
 PT sample with pair of ptsI primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 XX  
 PS Disclosure; SEQ ID NO 13; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of ptsI primers and probes to produce ptsI  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus ptsI gene.  
 XX  
 SQ Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 21; DB 12; Length 1379;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TGCATGATGGGTTATCTTCC 21  
 DB 209 TGCATGATGGGTTATCTTCC 189  
 RESULT 12  
 ADJ62004/C  
 ID ADJ62004 standard; DNA; 1379 BP.  
 XX  
 AC ADJ62004;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Group A streptococcus PtsI gene #11.  
 XX  
 KW Group A Streptococcus; GAS; gene; ds.  
 XX  
 OS Streptococcus sp.  
 XX  
 PN US2004014118-A1.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 XX  
 XX 20-FEB-2002; 2002US-00081923.  
 XX  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;

XX  
 DR WPI; 2004-224229/21.  
 XX  
 PT Detecting Group A Streptococcus in biological sample, involves contacting  
 PT sample with pair of ptsI primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 XX  
 XX Disclosure; SEQ ID NO 16; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of ptsI primers and probes to produce ptsI  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus ptsI gene.  
 XX  
 SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 21; DB 12; Length 1379;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TGCATGATGGGTTATCTTCC 21  
 DB 210 TGCATGATGGGTTATCTTCC 190  
 RESULT 13  
 ADJ61995/C  
 ID ADJ61995 standard; DNA; 1379 BP.  
 XX  
 AC ADJ61995;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Group A streptococcus PtsI gene #2.  
 XX  
 KW Group A Streptococcus; GAS; gene; ds.  
 XX  
 OS Streptococcus sp.  
 XX  
 PN US2004014118-A1.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 XX  
 XX 20-FEB-2002; 2002US-00081923.  
 XX  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;  
 XX  
 PD WPI; 2004-224229/21.  
 XX  
 PT Detecting Group A Streptococcus in biological sample, involves contacting  
 PT sample with pair of ptsI primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 XX  
 PS Disclosure; SEQ ID NO 7; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of ptsI primers and probes to produce ptsI  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample



CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus *ptai* gene.

XX  
 SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1379;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 |||||  
 Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 14  
 ADD68792/c  
 ID ADD68792 standard; DNA; 1384 BP.

XX  
 AC ADD68792;

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' *ptai* DNA - SEQ ID 8.

XX beta-haemolytic Group A Streptococcus; GAS;

KW Fluorescence resonance energy transfer; FRET; *ptai*; ds.

XX Streptococcus sp. 'group A'.

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

PI Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

XX Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.

XX Example 2; SEQ ID NO 8; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridising step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current  
 CC sequence is that of the Streptococcus sp. 'group A' *ptai* DNA (SEQ ID 8)  
 CC of the invention.

XX Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 |||||  
 Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 15

ADD68794/c

ID ADD68794 standard; DNA; 1384 BP.

XX  
 AC ADD68794;

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' *ptai* DNA - SEQ ID 10.

XX beta-haemolytic Group A Streptococcus; GAS;

KW fluorescence resonance energy transfer; FRET; *ptai*; ds.

XX Streptococcus sp. 'group A'.

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

XX Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.

XX Example 2; SEQ ID NO 10; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridising step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current  
 CC sequence is that of the Streptococcus sp. 'group A' *ptai* DNA (SEQ ID 10)  
 CC of the invention.

XX Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 |||||  
 Db 209 TGCATGTATGGGTTATCTTCC 189

RESULT 16

ADJ61998/c

ID ADJ61998 standard; DNA; 1384 BP.

XX  
 AC ADJ61998;

XX 06-MAY-2004 (first entry)

XX Group A streptococcus *ptai* gene #5.

XX Group A Streptococcus; GAS; gene; ds.

OS Streptococcus sp.  
 PN US2004014118-A1.  
 XX  
 XX 22-JAN-2004.  
 PD  
 XX  
 PF 19-JUN-2003; 2003US-00465205.  
 XX  
 PR 20-FEB-2002; 2002US-00081923.  
 XX  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 PA  
 XX Uhl JR, Cockerill FR;  
 PI  
 XX WPI; 2004-224229/21.  
 DR  
 XX  
 XX  
 PT Detecting Group A Streptococcus in biological sample, involves contacting  
 PT sample with pair of ptsI primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 PS Disclosure; SEQ ID NO 10; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of ptsI primers and probes to produce ptsI  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus ptsI gene.  
 XX  
 SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 21; DB 12; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TGCATGTATGGTTATCTTCC 21  
 |||||||  
 Db 209 TGCATGTATGGTTATCTTCC 189  
 |||||||  
 RESULT 17  
 ID ADJ61996/c  
 ADJ61996 standard; DNA; 1384 BP.  
 XX  
 AC ADJ61996;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DT  
 XX Group A streptococcus PtsI gene #3.  
 DE  
 XX Group A Streptococcus; GAS; gene; ds.  
 KW  
 XX Streptococcus sp.  
 OS  
 XX US2004014118-A1.  
 PN  
 XX  
 XX 22-JAN-2004.  
 PD  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 PF  
 XX 20-FEB-2002; 2002US-00081923.  
 PR  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 PA  
 XX Uhl JR, Cockerill FR;  
 PI  
 XX WPI; 2004-224229/21.  
 DR  
 XX  
 XX  
 PT Detecting Group A Streptococcus in biological sample, involves contacting

PT sample with pair of ptsI primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 XX Disclosure; SEQ ID NO 8; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of ptsI primers and probes to produce ptsI  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus ptsI gene.  
 XX  
 SQ Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 21; DB 12; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TGCATGTATGGTTATCTTCC 21  
 |||||||  
 Db 210 TGCATGTATGGTTATCTTCC 190  
 |||||||  
 RESULT 18  
 ID ADD68793/c  
 ADD68793 standard; DNA; 1385 BP.  
 XX  
 AC ADD68793;  
 XX  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX Streptococcus sp. 'group A' ptsI DNA - SEQ ID 9.  
 DE  
 XX beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; ptsI; ds.  
 XX  
 OS Streptococcus sp. 'group A'.  
 XX  
 XX US6593093-B1.  
 PN  
 XX 15-JUL-2003.  
 PD  
 XX 20-FEB-2002; 2002US-00081923.  
 PF  
 XX 20-FEB-2002; 2002US-00081923.  
 PR  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 PA  
 XX Uhl JR, Cockerill FR;  
 PI  
 XX WPI; 2003-828265/77.  
 DR  
 XX  
 XX  
 PT Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.  
 XX  
 XX Example 2; SEQ ID NO 9; 35pp; English.  
 PS  
 XX The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridising step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current

```
CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 9)
CC of the invention.
XX
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 19
ADJ61997/c
ID ADJ61997 standard; DNA; 1385 BP.
XX
AC ADJ61997;
XX
DT 06-MAY-2004 (first entry)
XX
DE Group A streptococcus PtsI gene #4.
XX
KW Group A Streptococcus; GAS; gene; ds.
XX
OS Streptococcus sp.
XX
PN US2004014118-A1.
XX
PD 22-JAN-2004.
XX
PF 19-JUN-2003; 2003US-00465205.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2004-224229/21.
XX
PT Detecting Group A Streptococcus in biological sample, involves contacting
PT sample with pair of ptai primers and probes and detecting fluorescence
PT resonance energy transfer indicating Group A Streptococcus in sample.
XX
PS Disclosure; SEQ ID NO 9; 36pp; English.
XX
CC The invention relates to a method of detecting the presence or absence of
CC Group A Streptococcus (GAS) in biological sample from individual, by
CC contacting sample with pair of ptai primers and probes to produce ptai
CC amplification product, where first, second probes are labelled with donor
CC and acceptor fluorescent moiety, and detecting presence or absence of
CC fluorescence resonance energy transfer, which indicates the presence or
CC absence of GAS in biological sample. The method is useful for detecting
CC the presence or absence of Group A Streptococcus in a biological sample
CC such as throat swabs, tissues and bodily fluids from an individual. The
CC present sequence represents a group A streptococcus ptai gene.
XX
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 20
ADD68795/c
ID ADD68795 standard; DNA; 1390 BP.
```

```
XX
AC ADD68795;
XX
DT 15-JAN-2004 (first entry)
XX
DE Streptococcus sp. 'group A' ptai DNA - SEQ ID 11.
XX
KW beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptai; ds.
XX
OS Streptococcus sp. 'group A'.
XX
PN US6593093-B1.
XX
PD 15-JUL-2003.
XX
PF 20-FEB-2002; 2002US-00081923.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
PS Example 2; SEQ ID NO 11; 35pp; English.
XX
CC The invention relates to a novel method for detecting the presence or
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC isolated from an individual by performing more than one cycling step. The
CC method involves an amplifying step comprising contacting the sample with
CC GAS primers, a subsequent hybridizing step comprising contacting the
CC amplified product with fluorescent-labelled probes and finally, detecting
CC fluorescence resonance energy transfer (FRET) where the presence or
CC absence of FRET indicates the presence or absence of GAS in the sample.
CC The method of the invention may be useful for detecting the presence or
CC absence of GAS in a biological sample from an individual. The current
CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 11)
CC of the invention.
XX
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1390;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 224 TGCATGTATGGGTTATCTTCC 204

RESULT 21
ADJ61999/c
ID ADJ61999 standard; DNA; 1390 BP.
XX
AC ADJ61999;
XX
DT 06-MAY-2004 (first entry)
XX
DE Group A streptococcus PtsI gene #6.
XX
KW Group A Streptococcus; GAS; gene; ds.
XX
OS Streptococcus sp.
XX
PN US2004014118-A1.
XX
```

```

PD 22-JAN-2004.
XX
XX
PF 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 11; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 12; Length 1390;
XX Best Local Similarity 100.0%; Pred. No. 2.1;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGTTCCTCC 21
DB 224 TGCATGTATGGTTCCTCC 204

RESULT 22
ADD68799/c
ID ADD68799 standard; DNA; 1393 BP.
XX
XX AC ADD68799;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptsI DNA - SEQ ID 15.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptsI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
XX performing cycling step comprising amplifying, hybridizing DNA from
XX sample by fluorescent probe, detecting for fluorescent resonance energy
XX transfer.
XX
XX Example 2; SEQ ID NO 15; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 15)
XX of the invention.
XX
XX Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 10; Length 1393;
XX Best Local Similarity 100.0%; Pred. No. 2.1;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGTTCCTCC 21
DB 210 TGCATGTATGGTTCCTCC 190

RESULT 23
ADDJ62003/c
ID ADDJ62003 standard; DNA; 1393 BP.
XX
XX AC ADDJ62003;
XX
XX 06-MAY-2004 (first entry)
XX
XX Group A streptococcus ptsI gene #10.
XX
XX Group A Streptococcus; GAS; gene; ds.
XX
XX Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 15; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 10; Length 1393;
XX Best Local Similarity 100.0%; Pred. No. 2.1;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGTTCCTCC 21
DB 210 TGCATGTATGGTTCCTCC 190

```

Query Match 100.0%; Score 21; DB 12; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;  
Matches 21; Conservative 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 24  
ADD6801/c  
ID ADD6801 standard; DNA; 1447 BP.  
XX  
AC ADD6801;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Streptococcus sp. 'group A' ptaI DNA - SEQ ID 17.  
XX  
KW beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptaI; ds.  
XX  
OS Streptococcus sp. 'group A'.  
XX  
PN US6593093-B1.  
XX  
PD 15-JUL-2003.  
XX  
PF 20-FEB-2002; 2002US-00081923.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2003-828265/77.  
XX

PT Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 17; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridizing step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 17)  
XX of the invention.

XX Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 25  
ADJ62005/c  
ID ADJ62005 standard; DNA; 1447 BP.

XX ADJ62005;  
AC 06-MAY-2004 (first entry)  
DT  
XX  
DE Group A streptococcus PtaI gene #12.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2004-224229/21.  
XX

PT Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptaI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.

XX Disclosure; SEQ ID NO 17; 36pp; English.

XX The invention relates to a method of detecting the presence or absence of  
XX Group A Streptococcus (GAS) in biological sample from individual, by  
XX contacting sample with pair of ptaI primers and probes to produce ptaI  
XX amplification product, where first, second probes are labelled with donor  
XX and acceptor fluorescent moiety, and detecting presence or absence of  
XX fluorescence resonance energy transfer, which indicates the presence or  
XX absence of GAS in biological sample. The method is useful for detecting  
XX the presence or absence of Group A Streptococcus in a biological sample  
XX such as throat swabs, tissues and bodily fluids from an individual. The  
XX present sequence represents a group A streptococcus ptaI gene.

XX Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 21; Conservative 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 26  
ABN68277/c  
ID ABN68277 standard; DNA; 1731 BP.  
XX  
AC ABN68277;  
XX

XX 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 4467.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX Streptococcus pyogenes.

XX WO200234771-A2.

XX 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27646.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
PT  
XX  
XX Claim 7; Page 3608; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus (I) may be prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
XX Sequence 1731 BP; 517 A; 358 C; 389 G; 467 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 2,2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 377 TGCATGTATGGGTTATCTTCC 357  
RESULT 27  
ABN68276/c  
ID ABN68276 standard; DNA; 1731 BP.  
XX  
AC ABN68276;  
XX  
XX 01-JUL-2002 (first entry)  
XX  
XX Streptococcus polynucleotide SEQ ID NO 4465.  
DE  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO200234771-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-GB004789.  
PF  
XX 27-OCT-2000; 2000GB-00026333.  
PR

PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27645.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
PT  
XX  
XX Claim 7; Page 3607; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC Streptococcus (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus (I) may be prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
XX Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 2,2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 377 TGCATGTATGGGTTATCTTCC 357  
RESULT 28  
ACA50572/c  
ID ACA50572 standard; DNA; 1734 BP.  
XX  
AC ACA50572;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #32229.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Streptococcus pyogenes.  
XX  
XX WO200277183-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
PR

```
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX PI WPI; 2003-029926/02.
XX DR P-PSDB; ABU46702.
XX DR
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 38442; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1734 BP; 519 A; 358 C; 389 G; 468 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 21; DB 8; Length 1734;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGCATGTATGGTTATCTTCC 21
Db 377 TGCATGTATGGTTATCTTCC 357
RESULT 29
ADV85417/c
ID ADV85417 standard; DNA; 1734 BP.
XX AC ADV85417;
XX
XX 24-FEB-2005 (first entry)
XX DE Streptococcus agalactiae DNA sequence, SEQ ID 6558.
XX KW Antibacterial; vaccine; bacterial infection; ds.
XX OS Streptococcus agalactiae.
XX PN WO200292819-A2.
XX
XX (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX PI WPI; 2003-029926/02.
XX DR P-PSDB; ABU46702.
XX DR
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 38442; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1734 BP; 519 A; 358 C; 389 G; 468 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 21; DB 8; Length 1734;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGCATGTATGGTTATCTTCC 21
Db 377 TGCATGTATGGTTATCTTCC 357
RESULT 29
ADV85417/c
ID ADV85417 standard; DNA; 1734 BP.
XX AC ADV85417;
XX
XX 24-FEB-2005 (first entry)
XX DE Streptococcus agalactiae DNA sequence, SEQ ID 6558.
XX KW Antibacterial; vaccine; bacterial infection; ds.
XX OS Streptococcus agalactiae.
XX PN WO200292819-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-1B003059.
XX
XX 26-APR-2001; 2001PR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 4; SEQ ID NO 6558; 439pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV81341-ADV85476) and
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
XX nucleotide sequences encode polypeptides of S. agalactiae involved in the
XX synthesis of amino acids, cell membranes, intermediate (central)
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,
XX regulatory functions, replication, transcription, translation, protein
XX transport, adaptation to atypical conditions, sensitivity to medicines
XX and/or analogues, functions related to transposons, biosynthesis of
XX cofactors, prosthetic groups and transporters, cell membrane proteins and
XX cellular machinery. (I) are useful for the detection and/or amplification
XX of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
XX useful for treatment of a bacterial S. agalactiae infection. The complete
XX genome of Streptococcus agalactiae is given in ADV81204. Note: The
XX present patent is an equivalent for the basic patent FR2824074A1, which
XX contains only 2344 sequences.
XX
XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 21; DB 13; Length 1734;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGCATGTATGGTTATCTTCC 21
Db 377 TGCATGTATGGTTATCTTCC 357
RESULT 30
ADD68806/c
ID ADD68806 standard; DNA; 1803 BP.
XX
XX ADD68806;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptaI DNA - SEQ ID 5.
XX DE beta-haemolytic Group A Streptococcus; GAS;
XX KW fluorescence resonance energy transfer; FRET; ptaI; ds; gene.
XX OS Streptococcus sp. 'group A'.
XX
XX Key Location/Qualifiers
XX CDS 1..1803
XX /tag= a
XX /partial
XX /product= "Streptococcus sp. 'group A' ptaI protein - SEQ
XX ID 6"
XX /note= "No stop codon; this CDS contains translation
XX exceptions"
```

```
XX PN US6593093-B1.
XX XX
XX PD 15-JUL-2003.
XX PF
XX PP 20-FEB-2002; 2002US-00081923.
XX PR
XX PX 20-FEB-2002; 2002US-00081923.
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;
XX PP WPI; 2003-828265/77.
XX PR P-PSDB; ADJ68790.
XX PX
XX PT Detecting presence/absence of group A Streptococcus in sample by
XX PP performing cycling step comprising amplifying, hybridizing DNA from
XX PT sample by fluorescent probe, detecting for fluorescent resonance energy
XX PT transfer.
XX PX
XX PP Example 2; SEQ ID NO 5; 35pp; English.
XX CC The invention relates to a novel method for detecting the presence or
XX CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX CC isolated from an individual by performing more than one cycling step. The
XX CC method involves an amplifying step comprising contacting the sample with
XX CC GAS primers, a subsequent hybridizing step comprising contacting the
XX CC amplified product with fluorescent-labelled probes and finally, detecting
XX CC fluorescence resonance energy transfer (FRET) where the presence or
XX CC absence of FRET indicates the presence or absence of GAS in the sample.
XX CC The method of the invention may be useful for detecting the presence or
XX CC absence of GAS in a biological sample from an individual. The current
XX CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 5)
XX CC of the invention.
XX SQ Sequence 1803 BP; 540 A; 371 C; 406 G; 486 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 10; Length 1803;
XX Best Local Similarity 100.0%; Pred. No. 2.2;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGGTATCTTCC 21
Db |||||
377 TGCATGTATGGGTATCTTCC 357

RESULT 31
ADJ61993/c
ID ADJ61993 standard; DNA; 1803 BP.
XX AC
XX ADJ61993;
XX XX
XX DT 06-MAY-2004 (first entry)
XX DE Group A streptococcus PtaI gene #1.
XX KW Group A Streptococcus; GAS; gene; ds.
XX OS Streptococcus sp.
XX PH Key Location/Qualifiers
XX FT 1..1803
XX FT /*tag= a
XX FT /product= "PtaI"
XX FT /note= "This sequence contains translation exceptions"
XX
XX PN US2004014118-A1.
XX XX
XX PD 22-JAN-2004.
XX PF
XX PP 19-JUN-2003; 2003US-00465205.
XX XX

PR 20-FEB-2002; 2002US-00081923.
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PA Uhl JR, Cockerill FR;
XX PI WPI; 2004-224229/21.
XX DR P-PSDB; ADJ61994.
XX
XX PT Detecting Group A Streptococcus in biological sample, involves contacting
XX PT sample with pair of ptai primers and probes and detecting fluorescence
XX PT resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX PS Disclosure; SEQ ID NO 5; 36pp; English.
XX
XX CC The invention relates to a method of detecting the presence or absence of
XX CC Group A Streptococcus (GAS) in biological sample from individual, by
XX CC contacting sample with pair of ptai primers and probes to produce ptai
XX CC amplification product, where first, second probes are labelled with donor
XX CC and acceptor fluorescent moiety, and detecting presence or absence of
XX CC fluorescence resonance energy transfer, which indicates the presence or
XX CC absence of GAS in biological sample. The method is useful for detecting
XX CC the presence or absence of Group A Streptococcus in a biological sample
XX CC such as throat swabs, tissues and bodily fluids from an individual. The
XX CC present sequence represents a group A streptococcus ptai gene.
XX SQ Sequence 1803 BP; 540 A; 371 C; 406 G; 486 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 12; Length 1803;
XX Best Local Similarity 100.0%; Pred. No. 2.2;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGGTATCTTCC 21
Db |||||
377 TGCATGTATGGGTATCTTCC 357

RESULT 32
ADV87723/c
ID ADV87723 standard; DNA; 29072 BP.
XX AC
XX ADV87723;
XX XX
XX DT 24-FEB-2005 (first entry)
XX DE Streptococcus agalactiae DNA sequence, SEQ ID 117.
XX KW Antibacterial; Vaccine; bacterial infection; ds.
XX OS Streptococcus agalactiae.
XX XX
XX PN FR2824074-A1.
XX XX
XX PD 31-OCT-2002.
XX XX
XX PF 26-APR-2001; 2001FR-00005642.
XX XX
XX PR 26-APR-2001; 2001FR-00005642.
XX XX
XX PA (INSP ) INST PASTEUR.
XX XX (CNRS ) CNRS CENT NAT RECH SCI.
XX XX
XX PI Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;
XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX XX WPI; 2004-101891/11.
XX
XX PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX PT and identification of therapeutic targets.
XX
XX PS Claim 1; SEQ ID NO 117; 2687pp; French.
XX XX
```



CC The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
CC agalactiae involved in the synthesis of amino acids, cell membranes,  
CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
CC phospholipid metabolism, nucleotide metabolism including purines,  
CC pyrimidines and/or nucleosides, regulatory functions, replication,  
CC transcription, translation, protein transport, adaptation to atypical  
CC conditions, sensitivity to medicines and/or analogues, functions related  
CC to transposons, biosynthesis of cofactors, prosthetic groups and  
CC transporters, cell membrane proteins and cellular machinery. (I) are  
CC useful for the detection and/or amplification of nucleic acids.  
CC Pharmaceutical composition comprising (I) or (II) are useful for  
CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 12272 TGCATGTATGGGTATCTTCC 12252

RESULT 33  
ADV78976/c  
ID ADV78976 standard; DNA; 29072 BP.  
XX  
AC ADV78976;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
DE  
XX Antibacterial; vaccine; bacterial infection; ds.  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO200292818-A2.  
XX  
XX 21-NOV-2002.  
XX  
XX 26-APR-2002; 2002WO-IB003059.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier P, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 439pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae  
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
XX nucleotide sequences encode polypeptides of S. agalactiae involved in the  
XX synthesis of amino acids, cell membranes, intermediate (central)  
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
XX regulatory functions, replication, transcription, translation, protein

CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 12272 TGCATGTATGGGTATCTTCC 12252

RESULT 34  
ABN71527\_07/c  
Continuation (8 of 22) of ABN71527 from base 700001 (Streptococcus polynucleotide SEQ 1  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527  
WP Fragment Name Begin End  
WP ABN71527\_00 1 110000  
WP ABN71527\_01 100001 210000  
WP ABN71527\_02 200001 310000  
WP ABN71527\_03 300001 410000  
WP ABN71527\_04 400001 510000  
WP ABN71527\_05 500001 610000  
WP ABN71527\_06 600001 710000  
WP ABN71527\_07 700001 810000  
WP ABN71527\_08 800001 910000  
WP ABN71527\_09 900001 1010000  
WP ABN71527\_10 1000001 1110000  
WP ABN71527\_11 1100001 1210000  
WP ABN71527\_12 1200001 1310000  
WP ABN71527\_13 1300001 1410000  
WP ABN71527\_14 1400001 1510000  
WP ABN71527\_15 1500001 1610000  
WP ABN71527\_16 1600001 1710000  
WP ABN71527\_17 1700001 1810000  
WP ABN71527\_18 1800001 1910000  
WP ABN71527\_19 1900001 2010000  
WP ABN71527\_20 2000001 2110000  
WP ABN71527\_21 2100001 2155561

Query Match 100.0%; Score 21; DB 6; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 91142 TGCATGTATGGGTATCTTCC 91122

RESULT 35  
ADV81204\_08/c  
Continuation (9 of 23) of ADV81204 from base 800001 (Streptococcus agalactiae complete  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204  
WP Fragment Name Begin End  
WP ADV81204\_00 1 110000  
WP ADV81204\_01 100001 210000  
WP ADV81204\_02 200001 310000  
WP ADV81204\_03 300001 410000  
WP ADV81204\_04 400001 510000  
WP ADV81204\_05 500001 610000  
WP ADV81204\_06 600001 710000  
WP ADV81204\_07 700001 810000  
WP ADV81204\_08 800001 910000  
WP ADV81204\_09 900001 1010000

WP ADV81204\_10 1110000  
 WP ADV81204\_11 1100001  
 WP ADV81204\_12 1210000  
 WP ADV81204\_13 1300001  
 WP ADV81204\_14 1400001  
 WP ADV81204\_15 1500001  
 WP ADV81204\_16 1600001  
 WP ADV81204\_17 1700001  
 WP ADV81204\_18 1800001  
 WP ADV81204\_19 1900001  
 WP ADV81204\_20 2000001  
 WP ADV81204\_21 2100001  
 WP ADV81204\_22 2200001  
 WP ADV81204\_23 2217924

Query Match 100.0%; Score 21; DB 13; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 3.6;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 Db 71477 TGCATGTATGGGTTATCTTCC 71457

RESULT 36  
 ID ABX06888/c  
 AC ABX06888; DNA; 402 BP.  
 XX  
 XX  
 XX 27-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 XX S. pneumoniae type 4 strain coding region #1176.  
 XX  
 XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 XX Streptococcus pneumoniae; type 4 strain.  
 OS  
 XX  
 XX WO200277021-A2.  
 XX  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Masignani V, Tettelin H, Fraser C;  
 PI  
 XX WPI: 2003-040579/03.  
 DR P-PSDB; ABU01600.  
 DR  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 XX Claim 6; SEQ ID NO 2351; 56pp; English.  
 PS  
 XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2489  
 CC identified coding region from the genomic sequence. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 402 BP; 128 A; 85 C; 96 G; 93 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 10; Length 402;  
 Best Local Similarity 95.2%; Pred. No. 12;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 Db 377 TGCATGTATGGGTTATCTTCC 357

RESULT 37  
 ID AAV65250/c  
 AC AAV65250; DNA; 962 BP.  
 XX  
 XX AAV65250;  
 XX  
 XX 24-DEC-1998 (first entry)  
 DT  
 XX  
 XX DNA encoding a S. pneumoniae PRP-protein PTS enzyme I.  
 DE  
 XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;  
 KW virulence; antibody; infection; detection; treatment; ss.  
 KW  
 XX Streptococcus pneumoniae.  
 OS  
 XX WO9826072-A1.  
 PN  
 XX 18-JUN-1998.  
 PD  
 XX 09-DEC-1997; 97WO-US022578.  
 PF  
 XX 13-DEC-1996; 96US-0036281P.  
 PR  
 XX (ELIL ) LILLY & CO ELI.  
 PA  
 XX Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;  
 PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR, Skatrud PL;  
 PI Smith MC, Solenberg PJ, Treadway PJ, Young Bellido ML;  
 XX  
 XX WPI: 1998-348529/30.  
 DR P-PSDB; AAW80678.  
 DR  
 XX Streptococcus pneumoniae nucleic acid sequences - used in DNA chips for  
 PT evaluating gene expression, and identification of virulence genes.  
 PT  
 XX Claim 1; Page 105; 333pp; English.  
 PS  
 XX This DNA sequence encodes a Streptococcus pneumoniae PRP-protein PTS  
 CC enzyme I. The invention provides DNA sequences (AAV65201 to AAV65304)  
 CC from the Streptococcus pneumoniae genome and corresponding protein  
 CC sequences (AAW80605 to AAW80728). A recombinant host containing a vector

CC comprising any of the above nucleic acids can be used for the recombinant  
CC expression of the protein sequences. The invention also provides a DNA  
CC chip having arrayed on it at least 15 base pair fragment of any one or  
CC more of these DNA sequences. The DNA chip can be used methods for  
CC evaluating gene expression in *S. pneumoniae* and for identifying virulence  
CC genes in *S. pneumoniae*. Antibodies that selectively bind to the above  
CC proteins or peptide fragments can be used to treat *S. pneumoniae*  
CC infection. The antibodies can also be used to detect *S. pneumoniae* cells  
XX  
SQ Sequence 962 BP; 312 A; 196 C; 224 G; 230 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 962;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21  
Db 519 TGCATGTATGGTTATCTTCC 499  
|||||

RESULT 38  
ADK44676/c  
ID ADK44676 standard; DNA; 1140 BP.

AC ADK44676;

XX 24-FEB-2005 (first entry)

DT Streptococcus pneumoniae gene, Seq ID No 1191.

DE db; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.

KW Streptococcus pneumoniae.

OS US6699703-B1.

PN 02-MAR-2004.

PD 26-MAY-2000; 2000US-00583110.

XX 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

PR 30-JUN-1998; 98US-00107433.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;

XX WPI; 2004-212399/20.

DR P-PSDB; ADK47337.

XX New nucleic acid molecules and polypeptides useful for diagnosing,  
PT preventing and treating pathological conditions resulting from bacterial  
PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
PT screening.

XX Disclosure; SEQ ID NO 1191; 301pp; English.

XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
CC and polypeptides. The nucleic acids and proteins are useful for  
CC diagnosing, preventing and treating pathological conditions resulting  
CC from bacterial infection, such as *S. pneumoniae* infection. These may also  
CC be used for drug screening procedures. The present sequence represents a  
CC Streptococcus pneumoniae nucleic acid of the invention. Note: The  
CC sequence data for this patent did not appear in the printed specification  
CC but was obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

XX Sequence 1140 BP; 355 A; 239 C; 259 G; 287 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 13; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21  
Db 377 TGCATGTATGGTTATCTTCC 357  
|||||

RESULT 39  
ADR92834/c

ID ADR92834 standard; DNA; 1287 BP.

XX ADR92834;

XX 16-DEC-2004 (first entry)

XX Novel *S. pneumoniae* DNA sequence, SEQ ID 1469.

XX Meningitis; bacteraemia; pneumonia; otitis media; da;  
KW bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2004-697205/68.

DR P-PSDB; ADR95437.

XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.

XX Disclosure; SEQ ID NO 1469; 151pp; English.

XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92850 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed *S.*  
CC pneumoniae nucleic acid sequences. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

XX Sequence 1287 BP; 396 A; 271 C; 295 G; 325 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 13; Length 1287;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 TGCATGTATGGGTATCTTCC 21
Db 431 TGCATGTATGGGTGCTTCC 411

RESULT 40
ID AEA56704/c
XX AEA56704 standard; DNA; 1287 BP.
AC AEA56704;
XX
XX
XX 25-AUG-2005 (first entry)
XX
XX Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:1469.
DE
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;
XX vaccine; gene; ds.
XX
XX Streptococcus pneumoniae.
XX
XX US2005136404-A1.
XX
XX 23-JUN-2005.
XX
XX 10-JUL-2003; 2003US-00617320.
XX
XX 02-JUL-1997; 97US-0051553P.
XX 12-MAY-1998; 98US-0085131P.
XX 30-JUN-1998; 98US-00107433.
XX
XX (DOUC/) DOUCETTE-STAMM L A.
XX (BUSH/) BUSH D.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2005-477576/48.
XX P-PSDB; AEA59307.
XX
XX New isolated nucleic acid molecules and encoded polypeptides useful for
XX diagnosing, preventing or treating bacterial infections, particularly
XX Streptococcus pneumoniae infection.
XX
XX Claim 1; SEQ ID NO 1469; 144pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule for detecting,
XX preventing or treating pathological conditions resulting from bacterial
XX infection. The isolated nucleic acid comprises: (a) any of the 2603
XX nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
XX encoding a Streptococcus pneumoniae polypeptide comprising any of the
XX 2603 amino acid sequences of AEA57839 to AEA6041; or (c) a nucleotide
XX sequence of at least 8 nucleotides in length, where the sequence is
XX hybridizable to a nucleic acid having any of the nucleotide sequences in
XX (a). Also described: (1) a recombinant expression vector comprising the
XX above nucleic acid operably linked to a transcription regulatory element;
XX (2) a cell comprising the recombinant expression vector; (3) producing an
XX S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
XX consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
XX treating a subject for S. pneumoniae infection; (6) a recombinant or
XX substantially pure preparation of an S. pneumoniae polypeptide or its
XX fragment, where the polypeptide is selected from AEA57839 to AEA6041;
XX (7) a vaccine composition for preventing or treating an S. pneumoniae
XX infection, comprising an amount of the above nucleic acid or polypeptide;
XX (8) detecting the presence of a Streptococcus nucleic acid in a sample;
XX (9) a computer readable medium having recorded the nucleotide sequences
XX of AEA55236 to AEA57838; (10) a computer based system for identifying
XX fragments of the Streptococcus genome of commercial importance. The
XX composition and methods are useful for diagnosing, preventing or treating
XX bacterial infections, particularly S. pneumoniae infection. The present
XX sequence represents a S. pneumoniae ORF nucleic acid sequence from the
XX present invention. Note - The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from the USPTO web site.
XX

SQ Sequence 1287 BP; 396 A; 271 C; 295 G; 325 T; 0 U; 0 Other;
Query Match 92.4%; Score 19.4; DB 14; Length 1287;
Best Local Similarity 95.2%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTATCTTCC 21
Db 431 TGCATGTATGGGTGCTTCC 411

RESULT 41
ID AAS55546/c
XX AAS55546 standard; DNA; 1734 BP.
AC AAS55546;
XX
XX 13-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae DNA for cellular proliferation protein #117.
DE
XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX antibacterial; drug design.
XX
XX Streptococcus pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009180.
XX
XX 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
XX 26-MAY-2000; 2000US-0207272P.
XX 23-OCT-2000; 2000US-0242578P.
XX 27-NOV-2000; 2000US-0253625P.
XX 22-DEC-2000; 2000US-0257931P.
XX 16-FEB-2001; 2000US-0269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlssen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX P-PSDB; AAU37687.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Claim 27; SEQ ID NO 9183; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence encodes an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1734 BP; 531 A; 365 C; 390 G; 448 T; 0 U; 0 Other;
```

Query Match 92.4%; Score 19.4; DB 4; Length 1734;  
 Best Local Similarity 95.2%; Pred. No. 14;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 |||||  
 Db 377 TGCATGTATGGGTTATCTTCC 357

RESULT 42  
 ACA49934/C  
 ID ACA49934 standard; DNA; 1734 BP.  
 XX  
 AC ACA49934;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #31591.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX drug design; gene.  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO20027183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 WPI: 2003-029926/02.  
 DR P-PSDB; ABU46064.  
 DR  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 37804; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1734 BP; 531 A; 365 C; 390 G; 448 T; 0 U; 0 Other;  
 SQ  
 Query Match 92.4%; Score 19.4; DB 8; Length 1734;  
 Best Local Similarity 95.2%; Pred. No. 14;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 |||||  
 Db 377 TGCATGTATGGGTTATCTTCC 357

RESULT 43  
 AAV52227/C  
 ID AAV52227 standard; DNA; 8195 BP.  
 XX  
 AC AAV52227;  
 XX  
 DT 23-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.

XX Streptococcus pneumoniae; *S. pneumoniae*; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX Streptococcus pneumoniae.  
 OS  
 XX WO9818931-A2.  
 PN  
 XX 07-MAY-1998.  
 PD  
 XX 30-OCT-1997; 97WO-US019588.  
 PF  
 XX 31-OCT-1996; 96US-0029960P.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
 PI Dougherty BA;  
 PI  
 XX WPI: 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 727-732; 1409pp; English.

XX The present invention describes a computer readable medium which has the  
 CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 CC 391, identifying members of the library which contain sequences that  
 CC hybridise to the target sequence and isolating the nucleic acid molecules  
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 CC homologous to amplification primers derived from the fragment of the *S.*  
 CC pneumoniae genome to prime the amplification and isolating the amplified

CC sequences. The computer readable medium can be used in a computer-based  
CC system for identifying fragments of the *S. pneumoniae* genome of  
CC commercial importance, or expression modulating fragments of the *S.*  
CC *pneumoniae* genome. Products from the present invention can be used in  
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
CC for *S. pneumoniae*

XX  
SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 8195;  
Best Local Similarity 95.2%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
DB 1384 TGCATGTATGGGTTGCTTCC 1364

## RESULT 44

ABSS56454\_09  
Continuation (10 of 22) of ABSS56454 from base 900001 (Streptococcus pneumoniae type 4  
WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000
WP	ABSS56454_10	1000001	1110000
WP	ABSS56454_11	1100001	1210000
WP	ABSS56454_12	1200001	1310000
WP	ABSS56454_13	1300001	1410000
WP	ABSS56454_14	1400001	1510000
WP	ABSS56454_15	1500001	1610000
WP	ABSS56454_16	1600001	1710000
WP	ABSS56454_17	1700001	1810000
WP	ABSS56454_18	1800001	1910000
WP	ABSS56454_19	1900001	2010000
WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598

Query Match 92.4%; Score 19.4; DB 10; Length 110000;  
Best Local Similarity 95.2%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
DB 108618 TGCATGTATGGGTTGCTTCC 108638

## RESULT 45

ABSS56454\_10  
Continuation (11 of 22) of ABSS56454 from base 1000001 (Streptococcus pneumoniae type 4  
WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000
WP	ABSS56454_10	1000001	1110000
WP	ABSS56454_11	1100001	1210000
WP	ABSS56454_12	1200001	1310000

WP	ABSS56454_13	1300001	1410000
WP	ABSS56454_14	1400001	1510000
WP	ABSS56454_15	1500001	1610000
WP	ABSS56454_16	1600001	1710000
WP	ABSS56454_17	1700001	1810000
WP	ABSS56454_18	1800001	1910000
WP	ABSS56454_19	1900001	2010000
WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598

Query Match 92.4%; Score 19.4; DB 10; Length 110000;  
Best Local Similarity 95.2%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
DB 8618 TGCATGTATGGGTTGCTTCC 8638

## RESULT 46

ACH68181  
ID ACH68181 standard; DNA; 553 BP.

XX	ACH68181;
XX	29-JUL-2004 (first entry)
XX	Human genome derived single exon probe #1376.
XX	Human; probe; ss; Gene expression; single exon probe; microarray;
XX	alternative splicing event; genomic alteration.
XX	Homo sapiens.
XX	US2003194704-A1.
XX	16-OCT-2003.
XX	03-APR-2002; 2002US-00029386.
XX	03-APR-2002; 2002US-00029386.
XX	(PENN/) PENN S G.
XX	(RANK/) RANK D R.
XX	(HANZ/) HANZEL D K.

PI Penn SG, Rank DR, Hanzel DK;  
XX WPI; 2004-119264/12.

XX New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.

XX Claim 15; SEQ ID NO 1376; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 688 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridises under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an

CC isolated antibody that binds specifically to a peptide cited above,  
 CC methods of selling and/or licensing single exon probes or microarrays to  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subcription, and a method of providing  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterising  
 CC alternative splicing events, in detecting and characterising gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704

XX Sequence 553 BP; 125 A; 160 C; 151 G; 117 T; 0 U; 0 Other;

Query Match 87.6%; Score 18.4; DB 12; Length 553;  
 Best Local Similarity 95.0%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTATCTTC 20  
 |||||  
 Db 213 TGCATGTATGGGTATCTTC 232

## RESULT 47

ADBI1789/c  
 ID ADBI1789 standard; DNA; 1830 BP.

XX AC ADBI1789;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Alloicoccus otitis antigenic protein encoding DNA SEQ ID NO:5273.  
 XX KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;  
 XX KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
 XX OS Alloicoccus otitis.

XX WO2003048304-A2.

XX PD 12-JUN-2003.

XX PF 25-NOV-2002; 2002WO-US036123.

XX PR 29-NOV-2001; 2001US-0333777P.

XX PR 18-NOV-2002; 2002US-0426742P.

XX PA (AMHP ) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX P-PSDB; ADBI1786.

XX New Alloicoccus otitis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 5273; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of  
 CC Alloicoccus otitis genomic DNA, which encodes an antigenic protein.  
 CC Alloicoccus otitis is a Gram-positive bacterium. Also described: (1)  
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (I), its

CC complement, degenerate variant or fragment; (3) a genetically engineered  
 CC host cell, transfected, transformed or infected with the vector of (2);  
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 CC composition comprising the polypeptide, its complement, biological  
 CC equivalent or fragment, or the polynucleotide that is comprised in the  
 CC expression vector; (6) a pharmaceutical composition comprising the  
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
 CC immunising against Alloicoccus otitis by administering to a host the  
 CC immunogen composition; (9) detecting and/or identifying Alloicoccus  
 CC otitis in the biological sample; (10) a kit comprising a container  
 CC containing the novel polynucleotide, its degenerate variant or fragment,  
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
 CC genetically engineered host cell under conditions suitable to produce the  
 CC polypeptide from the culture. (I) can be used in gene therapy. The  
 CC polynucleotides, polypeptides, antibodies and compositions of the present  
 CC invention can be used for treating and diagnosing diseases, drug  
 CC screening assays and monitoring of effects during drug clinical trials.  
 CC The polynucleotides are useful for expressing and detecting Alloicoccus  
 CC otitis. The present sequence encodes an Alloicoccus otitis antigen  
 CC protein from the present invention.

XX Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
 Best Local Similarity 90.5%; Pred. No. 89;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTATCTTC 21  
 |||||  
 Db 485 TGCATGTAGGGTTGTCTTC 465

## RESULT 48

ADBI1785/c

ID ADBI1785 standard; DNA; 1830 BP.

XX AC ADBI1785;

XX DT 20-NOV-2003 (first entry)

XX DE Alloicoccus otitis antigenic protein encoding DNA SEQ ID NO:5277.

XX KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;

XX KW gene therapy; Gram-positive bacterium; infection; gene; ds.

XX OS Alloicoccus otitis.

XX WO2003048304-A2.

XX PD 12-JUN-2003.

XX PF 25-NOV-2002; 2002WO-US036123.

XX PR 29-NOV-2001; 2001US-0333777P.

XX PR 18-NOV-2002; 2002US-0426742P.

XX PA (AMHP ) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX P-PSDB; ADBI1782.

XX New Alloicoccus otitis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 5277; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of  
 CC Alloicoccus otitis genomic DNA, which encodes an antigenic protein.  
 CC Alloicoccus otitis is a Gram-positive bacterium. Also described: (1)  
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (I), its

CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (1), its  
 CC complement, degenerate variant or fragment; (3) a genetically engineered  
 CC host cell, transfected, transformed or infected with the vector of (2);  
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 CC composition comprising the polypeptide, its complement, biological  
 CC equivalent or fragment, or the polynucleotide that is comprised in the  
 CC expression vector; (6) a pharmaceutical composition comprising the  
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
 CC immunising against *Alloiooccus* otitis by administering to a host the  
 CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*  
 CC otitis in the biological sample; (10) a kit comprising a container  
 CC containing the novel polynucleotide, its degenerate variant or fragment,  
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
 CC genetically engineered host cell under conditions suitable to produce the  
 CC polypeptide from the culture. (1) can be used in gene therapy. The  
 CC polynucleotides, polypeptides, antibodies and compositions of the present  
 CC invention can be used for treating and diagnosing diseases, drug  
 CC screening assays and monitoring of effects during drug clinical trials.  
 CC The polynucleotides are useful for expressing and detecting *Alloiooccus*  
 CC otitis. The present sequence encodes an *Alloiooccus* otitis antigen  
 CC protein from the present invention.

XX  
 SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
 Best Local Similarity 90.5%; Pred. No. 89;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTTAGGGTTATCTTCC 21  
 ||||| ||||| ||||| |||||  
 Db 485 TGCATGTTAGGGTTGCTTCC 465

RESULT 49  
 ADB11787/C  
 ID ADB11787 standard; DNA; 1830 BP.  
 AC ADB11787;  
 XX  
 XX 20-NOV-2003 (first entry)  
 XX  
 DE *Alloiooccus* otitis antigenic protein encoding DNA SEQ ID NO:5275.  
 XX  
 KW *Alloiooccus* otitis; antigenic protein; immunogenic; immunisation;  
 KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
 XX  
 OS *Alloiooccus* otitis.  
 XX  
 PN WO2003048304-A2.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 25-NOV-2002; 2002WO-US036123.  
 XX  
 PR 29-NOV-2001; 2001US-0333777P.  
 PR 18-NOV-2002; 2002US-0426742P.  
 XX  
 PA (AMHP ) WYETH HOLDINGS CORP.  
 XX  
 PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
 XX  
 DR WPI; 2003-505284/47.  
 DR P-PSDB; ADB11784.  
 XX  
 PT New *Alloiooccus* otitis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.  
 XX  
 PS Claim 7; SEQ ID NO 5275; 1019pp; English.  
 XX  
 CC The present invention describes an isolated polynucleotide (1) of

CC *Alloiooccus* otitis genomic DNA, which encodes an antigenic protein.  
 CC *Alloiooccus* otitis is a Gram-positive bacterium. Also described: (1)  
 CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (1), its  
 CC complement, degenerate variant or fragment; (3) a genetically engineered  
 CC host cell, transfected, transformed or infected with the vector of (2);  
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 CC composition comprising the polypeptide, its complement, biological  
 CC equivalent or fragment, or the polynucleotide that is comprised in the  
 CC expression vector; (6) a pharmaceutical composition comprising the  
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
 CC immunising against *Alloiooccus* otitis by administering to a host the  
 CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*  
 CC otitis in the biological sample; (10) a kit comprising a container  
 CC containing the novel polynucleotide, its degenerate variant or fragment,  
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
 CC genetically engineered host cell under conditions suitable to produce the  
 CC polypeptide from the culture. (1) can be used in gene therapy. The  
 CC polynucleotides, polypeptides, antibodies and compositions of the present  
 CC invention can be used for treating and diagnosing diseases, drug  
 CC screening assays and monitoring of effects during drug clinical trials.  
 CC The polynucleotides are useful for expressing and detecting *Alloiooccus*  
 CC otitis. The present sequence encodes an *Alloiooccus* otitis antigen  
 CC protein from the present invention.

XX  
 SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
 Best Local Similarity 90.5%; Pred. No. 89;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTTAGGGTTATCTTCC 21  
 ||||| ||||| ||||| |||||  
 Db 485 TGCATGTTAGGGTTGCTTCC 465

RESULT 50  
 ADB11783/C  
 ID ADB11783 standard; DNA; 1830 BP.  
 AC ADB11783;  
 XX  
 XX 20-NOV-2003 (first entry)  
 XX  
 DE *Alloiooccus* otitis antigenic protein encoding DNA SEQ ID NO:5279.  
 XX  
 KW *Alloiooccus* otitis; antigenic protein; immunogenic; immunisation;  
 KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
 XX  
 OS *Alloiooccus* otitis.  
 XX  
 PN WO2003048304-A2.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 25-NOV-2002; 2002WO-US036123.  
 XX  
 PR 29-NOV-2001; 2001US-0333777P.  
 PR 18-NOV-2002; 2002US-0426742P.  
 XX  
 PA (AMHP ) WYETH HOLDINGS CORP.  
 XX  
 PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
 XX  
 DR WPI; 2003-505284/47.  
 DR P-PSDB; ADB11780.  
 XX  
 PT New *Alloiooccus* otitis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.  
 XX  
 PS Claim 7; SEQ ID NO 5279; 1019pp; English.



XX The present invention describes an isolated polynucleotide (1) of  
CC Alloococcus otitidis genomic DNA, which encodes an antigenic protein.  
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
CC expression vector comprising the novel isolated polynucleotide (1), its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against Alloococcus otitidis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying Alloococcus  
CC otitidis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Alloococcus  
CC otitidis. The present sequence encodes an Alloococcus otitidis antigen  
CC protein from the present invention.  
XX  
SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;  
Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
Best Local Similarity 90.5%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
DB 485 TGCATGTAGGGTGTCTTCC 465  
RESULT 51  
ADR32245/c  
ID ADR32245 standard; cDNA; 2914 BP.  
XX ADR32245;  
AC ADR32245;  
XX 07-OCT-2004 (first entry)  
XX Human tumour suppressor BNO230 cDNA, SEQ ID NO:15.  
XX Tumour suppressor; cancer; breast; chromosome 16q24.3;  
KW loss of heterozygosity; LOH; transgenic animal; drug screening;  
KW diagnosis; prognosis; cytostatic; gene therapy; human; BNO230; FLJ31875;  
KW zinc finger domain; transcriptional regulation; gene; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH 333..1652  
FT /\*tag= a  
FT /product= "Tumour suppressor protein BNO230"  
XX WO200264780-A1.  
XX 22-AUG-2002.  
XX 12-FEB-2002; 2002WO-AU000138.  
XX 12-FEB-2001; 2001AU-00003052.  
PR 12-FEB-2001; 2001AU-00003053.  
PR 12-FEB-2001; 2001AU-00003054.  
XX (BION-) BIONOMICS LTD.  
XX

PI Callen DF, Powell JA, Kremmidiotis G, Gardner AB, Whitmore SA;  
XX WPI: 2002-667005/71.  
DR P-PSDB; ADR32246.  
XX Tumour suppressor genes at chromosome 16q24.3, useful for diagnosing,  
PT inhibiting or treating cancer or establishing the prognosis of a patient  
PT diagnosed with cancer, preferably breast cancer.  
XX Claim 11; SEQ ID NO 15; 129pp; English.  
XX The invention relates to 13 tumour suppressor genes involved in breast  
CC cancer, and their encoded proteins (ADR32231-ADR32256). The tumour  
CC suppressor proteins are PRDM7 (PR domain containing 7 protein), CDT1  
CC (CDT1 DNA replication factor), CHMP1 (charged multivesicular body protein),  
CC 1/chromatin modifying protein 1), BANP (BTG3 associated nuclear protein),  
CC BNO224, BNO36, BNO34, BNO208, BNO230 and BNO44. The genes encoding these  
CC proteins are located on chromosome 16q24.3, a region exhibiting loss of  
CC heterozygosity (LOH) in breast cancer and also in other carcinomas such  
CC as prostate cancers. The invention also relates to vectors and host cells  
CC comprising a tumour suppressor gene of the invention; antibodies against  
CC the tumour suppressor proteins; microarrays comprising fragments of the  
CC tumour suppressor genes; mutant forms of the tumour suppressor genes and  
CC proteins; methods of diagnosing, treating or inhibiting breast cancer; a  
CC genetically modified non-human animal in which a tumour suppressor gene  
CC is inactivated by knockout deletion; methods of screening for drugs which  
CC restore tumour suppressor activity; and pharmaceutical compositions  
CC comprising such a drug or a tumour suppressor gene, protein or antibody.  
CC The tumour suppressor genes and proteins of the invention may be used for  
CC diagnosing, inhibiting or treating cancer or for establishing the  
CC prognosis of a patient diagnosed with cancer, particularly breast cancer.  
CC The present sequence represents a cDNA encoding the tumour suppressor  
CC protein BNO230 (also known as FLJ31875). The BNO230 protein contains zinc  
CC finger domains, suggesting a role in DNA binding and/or protein complex  
CC interactions that regulate transcription.  
XX  
SQ Sequence 2914 BP; 821 A; 700 C; 728 G; 665 T; 0 U; 0 Other;  
Query Match 84.8%; Score 17.8; DB 7; Length 2914;  
Best Local Similarity 90.5%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
DB 1517 TGCATGTATGGGTATTATTC 1497  
RESULT 52  
ADAS3346/c  
ID ADAS3346 standard; cDNA; 2914 BP.  
XX ADAS3346;  
AC ADAS3346;  
XX 20-NOV-2003 (first entry)  
XX Human coding sequence, SEQ ID 914.  
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX Homo sapiens.  
XX EPI293569-A2.  
XX 19-MAR-2003.  
XX 21-MAR-2002; 2002EP-00006586.  
XX 14-SEP-2001; 2001JP-0028381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX (HELI-) HELIX RES INST.  
XX

PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hlo Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
DR WPI; 2003-395539/38.  
DR P-PSDB; ADA54985.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 1; SEQ ID NO 914; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX Sequence 2914 BP; 821 A; 700 C; 728 G; 665 T; 0 U; 0 Other;  
SQ  
Query Match 84.8%; Score 17.8; DB 10; Length 2914;  
Best Local Similarity 90.5%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 1517 TGCATGTATGGGTATCTTCC 1497  
RESULT 53  
ADBI2064\_13  
Continuation (14 of 18) of ADBI2064 from base 1300001 (Allootococcus otitis entire genome  
WP Sequence split into 18 fragments LOCUS ADBI2064 Accession Adbi2064  
WP Fragment Name Begin End  
WP ADBI2064\_00 1 110000  
WP ADBI2064\_01 100001 210000  
WP ADBI2064\_02 200001 310000  
WP ADBI2064\_03 300001 410000  
WP ADBI2064\_04 400001 510000  
WP ADBI2064\_05 500001 610000  
WP ADBI2064\_06 600001 710000  
WP ADBI2064\_07 700001 810000  
WP ADBI2064\_08 800001 910000  
WP ADBI2064\_09 900001 1010000  
WP ADBI2064\_10 1000001 1110000  
WP ADBI2064\_11 1100001 1210000  
WP ADBI2064\_12 1200001 1310000  
WP ADBI2064\_13 1300001 1410000  
WP ADBI2064\_14 1400001 1510000  
WP ADBI2064\_15 1500001 1610000  
WP ADBI2064\_16 1600001 1710000  
WP ADBI2064\_17 1700001 1754382  
Query Match 84.8%; Score 17.8; DB 9; Length 110000;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 94003 TGCATGTATGGGTATCTTCC 94023  
RESULT 54  
ABK74374  
ID ABK74374 standard; DNA; 412 BP.  
XX  
AC ABK74374;  
XX  
DT 13-AUG-2002 (first entry)  
XX

DE Bacillus licheniformis genomic sequence tag (GST) #1665.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031437.  
XX  
XX 06-OCT-2000; 2000US-00680598.  
PR 27-MAR-2001; 2001US-0279526P.  
XX  
XX (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
XX WPI; 2002-416684/44.  
DR  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second Bacillus  
PT cells, by using substrate containing Bacillus genomic sequenced tag  
PT array.  
XX  
PS Claim 4; SEQ ID NO 1665; 200pp; English.  
XX  
CC The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions, have follow  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 412 BP; 137 A; 74 C; 104 G; 97 T; 0 U; 0 Other;  
Query Match 80.0%; Score 16.8; DB 6; Length 412;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GCATGTATGGGTATCTTCC 21  
|||||  
Db 78 GCATGTATGGGTATCTGCC 97  
RESULT 55  
AAX30475/c  
ID AAX30475 standard; DNA; 1464 BP.  
XX  
AC AAX30475;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE H. pylori secreted protein ORF hp6p10606\_23493756\_c1\_21.

XX Vaccine; probe; diagnostic; ORF; cell envelope protein; secreted protein;  
 KW cellular protein; ds.

XX Helicobacter pylori.

XX WO9818323-A1.

XX 07-MAY-1998.

XX 28-OCT-1997; 97WO-US019575.

XX 28-OCT-1996; 96US-00739150.

XX 06-DEC-1986; 96US-00759739.

XX 14-JUL-1997; 97US-00891928.

XX (ASTR ) ASTRA AB.

XX Smith D, Alm RA;

XX WPI; 1998-271811/24.

XX P-PSDB; AAY11008.

XX Helicobacter pylori nucleic acids and proteins - used to develop products  
 for the detection, prevention and treatment of H. pylori infections.

XX Claim 3, 4; Page 137; 279pp; English.

XX Recombinant or substantially pure preparations of H. pylori polypeptides  
 are disclosed, together with the nucleic acids encoding them. In all, 73  
 ORFs are shown. The proteins are variously cell envelope proteins,  
 secreted proteins or other cellular proteins. Vaccines containing the  
 nucleic acids or proteins are claimed, as are probes containing at least  
 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
 for treating or reducing the risk of H. pylori infections, and the probes  
 can be used diagnostically for detecting the presence of Helicobacter in  
 a sample. The products are also of use in screening for compounds having  
 the ability to interfere with the H. pylori life cycle or to inhibit H.  
 pylori infection

SQ Sequence 1464 BP; 545 A; 273 C; 276 G; 370 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 1464;

Best Local Similarity 90.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 2;

Qy 1 TGCATGTATGGTTATCTTC 20

Db 1370 TGCATGTATGGTTTCTTC 1351

RESULT 56

AAAT68140/c

ID AAAT68140 standard; DNA; 1554 BP.

XX AAAT68140;

XX 18-JUL-1997 (first entry)

DE H. pylori inner membrane protein ORF 13ep12003orf20.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT 1. .1554

FT /\*tag= a

FT /note= "no stop codon given"

PN WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.

XX 01-APR-1996; 95US-00630405.

XX (ASTR ) ASTRA AB.

XX Smith D, Berglinth OT, Mellgaard BL;

XX WPI; 1997-052306/05.

XX P-PSDB; AAW20887.

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 useful for vaccines to treat or prevent H. pylori infection, and to  
 detect Helicobacter.

XX Claim 1; Page 924; 1481pp; English.

XX This sequence encodes a H. pylori inner membrane protein. The protein may  
 be used in a vaccine to prevent or treat H. pylori infection or to  
 identify H. pylori polypeptide binding compounds, useful as potential H.  
 pylori life cycle activators or inhibitors. The genomic sequence of H.  
 pylori (ATCC 55679) was determined from overlapping contigs generated by  
 mechanically shearing the bacterial DNA. The sequences were analysed for  
 ORF of at least 180 nucleotides, and the predicted coding regions defined  
 by computer evaluation. To identify likely H. pylori antigens for vaccine  
 development, the amino acid sequences predicted from various ORF were  
 analysed for significant homology to other known or exported membrane  
 proteins. Having identified and determined the sequences of interest,  
 particular regions can be isolated from H. pylori by PCR amplification  
 for recombinant polypeptide production, e.g. in E. coli hosts

SQ Sequence 1554 BP; 571 A; 281 C; 290 G; 412 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 1554;

Best Local Similarity 90.0%; Pred. No. 2.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 20

Db 1463 TGCATGTATGGTTTCTTC 1444

RESULT 57

ADE91329/c

ID ADE91329 standard; DNA; 2547 BP.

XX ADE91329;

XX 12-FEB-2004 (first entry)

DE Maize glutamine synthetase (GS1-2) promoter.

XX Maize; cytoplasmic glutamine synthetase promoter; GS1-2 promoter;  
 KW transgene expression; plant breeding; corn; transgenic; ds.

XX Zea mays.

XX US2003140364-A1.

XX 24-JUL-2003.

XX 20-NOV-2001; 2001US-00989739.

XX 20-NOV-2001; 2001US-00989739.

XX (HINC/) HINCHEY B.

XX (SONG/) SONG H.

```
PI Hinchey B, Song H;
XX
XX WPI; 2004-020397/02.
XX
XX Novel isolated nucleic acid sequence comprising cytoplasmic glutamine
PT synthetase promoter useful for improving feed or food value, improving
PT processing of corn and products resulting from processing.
XX
XX Claim 11; Fig 2; 62pp; English.
XX
XX The present invention relates to the isolation of maize cytoplasmic
CC glutamine synthetase (GS1-2) promoter. Also disclosed are methods for the
CC expression of transgenes in plants, stably transformed with a selected
CC DNA comprising a maize GS1-2 promoter. The GS1-2 promoter and the methods
CC are useful for plant breeding. The transgenic plant is as a
CC monocytyledonous plant (wheat, maize, rice, oat, barley, turfgrass,
CC sorghum, millet and sugarcane) or dicotyledonous plant (tobacco, tomato,
CC potato, soybean, cotton, canola, alfalfa, sunflower, and cotton).
CC Preferably, the monocytyledonous or dicotyledonous plant is maize or
CC soybean plant, respectively. The promoter is useful for directing the
CC expression of a selected coding region which encodes a particular protein
CC or polypeptide product, improving feed or food value, improving
CC processing of corn and for improving the value of the product resulting
CC from the processing. The promoter is efficient in expressing transgenes
CC in plants. The present sequence represents maize GS1-2 promoter.
XX
XX Sequence 2547 BP; 691 A; 602 C; 574 G; 680 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16.8; DB 12; Length 2547;
Best Local Similarity 90.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
DB 1824 TGCATGTCTGGGCTATCTTC 1805

RESULT 58
ABV72307/c
ID ABV72307 standard; DNA; 2953 BP.
AC
XX ABV72307;
XX
XX 13-MAY-2003 (first entry)
XX
XX Nucleotide sequence of fructosyl amino acid oxidase.
DE
XX Fructosyl amino acid oxidase; enzyme; gene; ss.
KW
XX Aspergillus oryzae.
OS
XX
XX Key Location/Qualifiers
FH 1038..2551
FT CDS /*tag= a
FT /*product= "fructosyl amino acid oxidase"
FT /*note= "contains introns"
FT exon 1038..1086
FT /*tag= b
FT /*number= 1
FT intron 1087..1224
FT /*tag= c
FT /*number= 1
FT exon 1225..2213
FT /*tag= d
FT /*number= 1
FT intron 2214..2269
FT /*tag= e
FT /*number= 2
FT exon 2270..2551
FT /*tag= f
FT /*number= 3
XX
XX JP2002218982-A.

XX 06-AUG-2002.
XX
XX 25-JAN-2001; 2001JP-00017640.
XX
XX 25-JAN-2001; 2001JP-00017640.
XX
XX (DOKU-) DOKURITSU GYOSEI HOJIN SHURUI SOGO KENKY.
XX (GEKK-) GEKKEIKAN KK.
XX
XX WPI; 2002-639259/69.
XX P-PSDB; ABB78358.
XX
XX Novel fructosyl amino acid oxidase protein for food drugs and cosmetics.
XX
XX Claim 2; Page 7-8; 15pp; Japanese.
XX
XX The present sequence encodes a fructosyl amino acid oxidase polypeptide.
XX The enzyme is useful for foods, drugs and cosmetics
XX
XX Sequence 2953 BP; 763 A; 749 C; 737 G; 704 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16.8; DB 6; Length 2953;
Best Local Similarity 90.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
DB 1441 TGCATGTATCGGTATCTTC 1422

RESULT 59
ADZ12700/c
ID ADZ12700 standard; DNA; 48179 BP.
XX
XX AC ADZ12700;
XX
XX 16-JUN-2005 (first entry)
XX
XX Murine cancer-associated genomic DNA #18.
DE
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW cytostatic; gene; ds.
XX
XX Mus sp.
OS
XX WO2005031001-A2.
PN
XX 07-APR-2005.
XX
XX 23-SEP-2004; 2004WO-US031617.
XX
XX 23-SEP-2003; 2003US-00669920.
XX (CHIR ) CHIRON CORP.
XX
XX Morris DW, Malandro MS;
PI WPI; 2005-273395/28.
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
XX
XX Disclosure; SEQ ID NO 220; 198pp; English.
XX
XX The invention relates to a nucleic acid array for detecting a cancer
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.
XX The invention also relates to a peptide array comprising two or more
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound
XX that binds to a polypeptide, an isolated antibody or its fragment which
XX binds to a polypeptide, which is prepared by immunizing a host animal
XX with a composition comprising the polypeptide or its antigen binding
XX fragment and collecting cells from the host expressing antibodies against
```

CC the antigen or its antigen binding fragment, a composition comprising the  
 CC antibody and a carrier, a method of screening for anticancer activity, a  
 CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
 CC method of treating cancer and a method of inhibiting expression of a CA  
 CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
 CC nucleic acids. The antibody is useful for detecting the presence or  
 CC absence of cancer cells in an individual which involves contacting cells  
 CC from the individual with the antibody and detecting a complex of a CA  
 CC protein from the cancer cells and the antibody, where the detection of  
 CC the complex correlates with the presence of cancer cells in the  
 CC individual. The composition is useful for inhibiting growth of cancer  
 CC cells in an individual or for delivering a therapeutic agent to cancer  
 CC cells in an individual. The invention is also useful for diagnosing  
 CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
 CC a cell. This sequence represents murine cancer-associated genomic DNA of  
 CC the invention.

XX  
 SQ Sequence 48179 BP; 13764 A; 10405 C; 10337 G; 13482 T; 0 U; 191 Other;

Query Match 80.0%; Score 16.8; DB 14; Length 48179;  
 Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21  
 ||||| ||||| ||||| |||||  
 Db 34371 GCATGTGTGGTTATCTTCC 34352

## RESULT 60

ADA02873/c

ID ADA02873 standard; DNA; 51281 BP.

AC ADA02873;

XX

DT 06-NOV-2003 (first entry)

XX

DE Mouse Dpt carcinoma associated gene, SEQ ID NO:1391.

XX

XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 XX gene; ds.

OS Mus sp.

XX W02003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

XX 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.

XX Claim 1; SEQ ID NO 1391; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a

CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 51281 BP; 13608 A; 10106 C; 10141 G; 12908 T; 0 U; 4518 Other;

Query Match 80.0%; Score 16.8; DB 9; Length 51281;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21

||||| ||||| ||||| |||||

Db 37474 GCATGTGTGGTTATCTTCC 37455

Search completed: January 27, 2006, 22:15:20  
 Job time : 201.315 secs

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 1701.94 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattatgggtatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_est9.\*

10: gb\_est10.\*

11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.4	87.6	815	8	DR937894 EST112943
C 2	17.8	84.8	443	5	BX556085 BX556085
C 3	17.8	84.8	456	1	A1163927 A051P45U
C 4	17.8	84.8	483	2	BG660206 TGESTz297
C 5	17.8	84.8	560	9	BH327941 CH230-117
C 6	17.8	84.8	619	11	DE064462 Oryzias 1
C 7	17.8	84.8	637	11	DE067970 Oryzias 1
C 8	17.8	84.8	714	3	BM645185 170006873
C 9	17.8	84.8	749	8	DN793467 Pm Admus
C 10	17.8	84.8	847	10	DU058511 90280 Tom
C 11	17.8	84.8	933	10	CZ954709 268426 To
C 12	17.8	84.8	949	2	BE913346 601665556
C 13	17.8	84.8	952	9	CC131506 NDL.45A4.
C 14	17.8	84.8	1035	10	CZ949984 263593 To
C 15	17.4	82.9	221	1	BB061307 BB061307
C 16	17.4	82.9	397	1	AA489071 UI-M-BH3-
C 17	17.4	82.9	403	4	AK219962 Mus muscu
C 18	17.4	82.9	416	6	CB799737 AMGNNUC:N
C 19	17.4	82.9	429	1	AA492409 UI-M-BH3-
C 20	17.4	82.9	459	6	CB738293 AMGNNUC:N
C 21	17.4	82.9	481	6	CA895102 B0189H07-
C 22	17.4	82.9	553	6	CA888060 B0142D05-

23	17.4	82.9	561	6	CA637985
24	17.4	82.9	581	5	BQ571868
25	17.4	82.9	589	5	BQ571429
26	17.4	82.9	592	10	CE495411 tigr-g88-
27	17.4	82.9	594	9	BZ300676 K00856.p1
28	17.4	82.9	596	9	CE104077 tigr-g88-
29	17.4	82.9	600	2	BE336403 601088981
30	17.4	82.9	643	9	CE133541 tigr-g88-
31	17.4	82.9	685	2	BE533917 601234117
32	17.4	82.9	691	6	CA318453 UI-M-FW0-
33	17.4	82.9	697	6	CF749756 UI-M-HJ0-
34	17.4	82.9	702	6	CA511105 UI-R-FJ0-
35	17.4	82.9	707	6	CB519329 UI-M-GH0-
36	17.4	82.9	714	8	CV805733 AGENCOURT
37	17.4	82.9	721	7	CK364526 AGENCOURT
38	17.4	82.9	767	6	CB245585 UI-M-FY0-
39	17.4	82.9	768	7	CO428850 UI-M-HX0-
40	17.4	82.9	777	6	CD349045 UI-M-FY0-
41	17.4	82.9	787	8	CK563730 UI-M-FY0-
42	17.4	82.9	793	6	CB521799 UI-M-GH0-
43	17.4	82.9	797	6	CF737017 UI-M-HD0-
44	17.4	82.9	802	6	CB290090 UI-M-FI0-
45	17.4	82.9	804	3	BQ180689 UI-M-EX0-
46	17.4	82.9	806	6	CF726722 UI-M-HB0-
47	17.4	82.9	814	5	BQ444317 UI-M-EX0-
48	17.4	82.9	840	3	B1685030 603310009
49	17.4	82.9	845	5	BU613524 UI-M-EM0-
50	17.4	82.9	961	11	CNS04600 Tetraodon
51	17.4	82.9	995	2	BG519143 602578081
52	17.4	82.9	1032	10	CL116220 1581-63M2
53	17.4	82.9	3610	4	AK029006 Mus muscu
54	17.4	82.9	3626	4	AK032373 Mus muscu
55	17.4	82.9	3732	4	AK083584 Mus muscu
56	17	81.0	668	10	CG735819 ZMMBB032
57	17	81.0	738	7	CN212867 26396 Sub
58	16.8	80.0	109	5	BW810341 BW810341
59	16.8	80.0	226	2	BB303972 BB303972
60	16.8	80.0	243	7	CN900771 010717ABB

#### ALIGNMENTS

RESULT 1  
DR937894/c 815 bp mRNA linear EST 02-AUG-2005  
LOCUS EST1129433 Aquilegia cDNA library Aquilegia formosa x Aquilegia  
DEFINITION pubescens cDNA clone COLPK16, mRNA sequence.  
ACCESSION DR937894.1 GI:71707257  
VERSION DR937894  
KEYWORDS EST  
SOURCE Aquilegia formosa x Aquilegia pubescens  
ORGANISM Aquilegia formosa x Aquilegia pubescens  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
Ranunculaceae; Aquilegia.  
REFERENCE 1 (bases 1 to 815)  
AUTHORS Hodges, S.A., Rensink, W., Buell, C.R., Borevitz, J., Kramer, E.,  
Nordborg, M. and Tomkins, J.  
TITLE Generation of ESTs from Aquilegia  
JOURNAL Unpublished (2005)  
COMMENT Other ESTs: EST1129434  
Contact: Scott Hodges  
Department of Ecology, Evolution and Marine Biology  
University of California, Santa Barbara  
Santa Barbara, CA 93106, USA  
Tel: 805 893 7813  
Fax: 805 893 4724  
Email: hodges@lifesci.ucsb.edu  
Seq primer: TTTTTCCTTTTTCCTTTTTCCTTTT (where N = A, G & C).  
Location/Qualifiers  
1. .815  
/organism="Aquilegia formosa x Aquilegia pubescens"

```

/mol_type="mRNA"
/db_xref="taxon:338618"
/clone="ColPK16"
/tissue_type="mixed shoot and floral apical meristems,
flower buds, leaves and roots"
/lab_host="DH10B T1 (T1 and T5 phage resistance)"
/clone_lib="Aquillegia cDNA library"
/note="Vector: pCMV SPORT6.1; Site 1: EcoRI; Site 2: NotI;
F2, F3, and F4 lines of Aquilegia formosa X A. pubescens
were grown from seed in greenhouses at UC Santa Barbara.
From these plants three sets of tissue were collected: 1)
Small flower buds (<10 mm) and very young inflorescences
(T1 & 29# by weight respectively), 2) Medium (7-20 mm) and
large (at or near anthesis) flower buds (65 & 35# by
weight respectively) and 3) Shoot apical meristems. A
fourth set of tissue was collected from plants of A.
formosa. These plants were grown from seed in sand at
approximately 1 month root tissue and leaf tissue of
various developmental stages were collected (84 & 16# by
weight respectively). Total RNA was extracted from each
set of tissue and pooled in the following proportions:
1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
total RNA, mRNA was extracted and enriched for full-length
messages and then normalized with proprietary methods by
Invitrogen."

```

## ORIGIN

```

Query Match      87.6%; Score 18.4; DB 8; Length 815;
Best Local Similarity 95.0%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 TGCATGTATGGTATCTTC 20
      ||||| ||||| ||||| |||||
DB      278 TGCATGTACGGTATCTTC 259

```

## RESULT 2

```

BX556085
LOCUS      BX556085 Glossina morsitans morsitans adult infected gut Glossina
DEFINITION morsitans morsitans cDNA clone Tse24g11_glc, mRNA sequence.
ACCESSION      BX556085.1 GI:33380027
VERSION
KEYWORDS
SOURCE
ORGANISM

```

## REFERENCE

```

AUTHORS      Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M.,
Hamilton, J., Soares, M.B., Bonaldo, M.F., Lehane, S. and Hall, N.
TITLE      Adult midgut expressed sequence tags from the tsetse fly Glossina
morsitans morsitans and expression analysis of putative immune
response genes

```

```

JOURNAL
PUBMED      14519198
COMMENT      Genome Biol. 4 (10), R63 (2003)

```

```

Contact: Hall N
Pathogen Sequencing Unit
The Sanger Institute The Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SA, UK
Request for clones, please contact: Mike Lehane
Prof. M.J. Lehane
School of Biological Sciences,
University of Wales,
Bangor LL57 2UW
All clones with suffix q1c are reverse primer reads starting at 5'
end of the cDNA all p1c reads are from
the 3' end.

```

## FEATURES

```

source
1. .443
/organisms="Glossina morsitans morsitans"
/mol_type="mRNA"

```

```

/sub_species="morsitans"
/db_xref="taxon:37546"
/clone="Tse24g11_glc"
/tissue_type="adult infected gut"
/clone_lib="Glossina morsitans morsitans adult infected
gut"
/note="Country: Zimbabwe; EST from adult gut infected with
T.brucei"

```

## ORIGIN

```

Query Match      84.8%; Score 17.8; DB 5; Length 443;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TCATGTATGGTATCTTC 21
      ||||| ||||| ||||| |||||
DB      66 TCCATGTATGGTATCTTC 86

```

## RESULT 3

```

Ail63927/c
LOCUS      Ail63927/c 456 bp mRNA linear EST 03-DEC-1998
DEFINITION A051P45U Hybrid aspen plasmid library Populus tremula x Populus
tremuloides cDNA 5', mRNA sequence.
ACCESSION      Ail63927
VERSION
KEYWORDS
SOURCE
ORGANISM

```

## REFERENCE

```

AUTHORS      Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W.,
Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.
TITLE      Gene discovery in the wood-forming tissues of poplar: Analysis of
5,692 expressed sequence tags

```

```

JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
PUBMED      9789088
COMMENT      Contact: Sterky F

```

```

Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGTGCTGCAGGCG
BACKWARD: GCTCCGGCTCGTATGTGTGTG
Seq primer: CGTTGTAACGACGCCAG
High quality sequence stop: 456.

```

```

Location/Qualifiers
1. .456
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"

```

```

/db_xref="taxon:47664"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/clone_lib="Hybrid aspen plasmid library"
/note="Vector: pBluescript SK; Site 1: SalI; Site 2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

```

## FEATURES

```

source

```

## ORIGIN

```

Query Match      84.8%; Score 17.8; DB 1; Length 456;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;

```



```

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
    ||||| ||||| ||||| |||||
Db 419 TGCATGTAGGTTACCTCC 399

RESULT 4
BG660206 483 bp mRNA linear EST 11-MAY-2001
LOCUS TGEStz97h11.y1 TgVEG118 Tachyzoite cDNA Library Toxoplasma gondii
DEFINITION cDNA clone TGEStz97h11.y1 5' similar to TR:Q9VHS8 Q9VHS8 CG7483
PROTEIN ; mRNA sequence.
ACCESSION BG660206
VERSION BG660206.1 GI:13803341
KEYWORDS EST.
SOURCE Toxoplasma gondii
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 483)
Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajioka,J.A., White,M.,
Clifton,S., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M.,
Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M.,
Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I.,
Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxest@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 369.
Location/Qualifiers
1 .483
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="VEG"
/db_xref="taxon:5811"
/clones="TGEStz97h11.y1"
/dev_stage="Tachyzoite"
/lab_host="DH10B"
/clone_lib="TgVEG118 Tachyzoite cDNA Library"
/notes="Vector: pBluescript SK; Site1: EcoRI; Site_2:
XhoI; This library was constructed by Keliang Tang, Robert
Cole, and L. David Sibley at Washington University. CDNAS
were synthesized from poly(A)+ RNA by o1idd(T) priming,
size-selected and directionally cloned into the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemids and rescued in SOLR cells. The
plasmid library was recovered from the SOLR cells and
transformed in mass into DH10B (GeneHog, Research
Genetics, Inc.) for sequencing. WARNING: This library may
contain a small percentage contaminants from human
fibroblast cells."

ORIGIN
Query Match 84.8%; Score 17.8; DB 2; Length 483;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
    ||||| ||||| ||||| |||||
Db 387 TGCATGTCTGGGTTCTTCC 407

RESULT 5

```

```

BH327941/c
LOCUS BH327941
DEFINITION CH230-117B21.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH327941
VERSION BH327941.1 GI:17258655
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 560)
Zhao,S., Shetty,J., Shatsman,S., Tsagsay,G., Geer,K.,
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Other GSSs: CH230-117B21.TV
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering.information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 117 row: B column: 21
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1 .560
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SnHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-117B21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/notes="Vector: pTARBAC2.1; Site.1: EcoRI; Site.2: EcoRI;
CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 560;
Best Local Similarity 90.5%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
    ||||| ||||| ||||| |||||
Db 546 TGTATGCTGGTTATCTTCC 526

RESULT 6
DE064462/c
LOCUS DE064462
DEFINITION Oryzias latipes DNA, clone: olal-111B07.R, genomic survey sequence.
ACCESSION DE064462
VERSION DE064462.1 GI:62576006
KEYWORDS GSS.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1

```

```

AUTHORS      Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.
TITLE        BAC end sequences of Olal Oryzias latipes Library
JOURNAL      Published Only in Database (2005)
REFERENCE    2 (bases 1 to 619)
AUTHORS      Fujiyama, A.
TITLE        Direct Submission
JOURNAL      Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
              (E-mail: afujiyam@gsc.riken.jp, URL: http://stt.gsc.riken.jp/,
              Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)
              This work was done in collaboration with Takeda, H. (1), Naruse, K.
              (2)
              and Narita, T. (3)
              (1) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: htakeda.s.u-tokyo.ac.jp
              (2) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: naruse.s.u-tokyo.ac.jp
              (3) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: tanarita.s.u-tokyo.ac.jp
              PRIMERS
              Sequencing : Forward
LIBRARY      Vector      : pKS145
              R.Site 1   : SacI
              L.Site 2   : SacI
              Location/Qualifiers
FEATURES     source
              1..619
              /organism="Oryzias latipes"
              /mol_type="genomic DNA"
              /db_xref="taxon:8090"
              /clone="olal-164P16.F"
              /sex="male"
              /cell_type="whole body"
              /clone_lib="BAC end sequences of Olal Oryzias latipes
              library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 11; Length 619;
Best Local Similarity 90.5%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTCATGGTTATCTTCC 21
    |||||
Db 518 TGCATGTCGCGTTTCTTCC 498

RESULT 7
LOCUS      DE067970/c
DEFINITION Oryzias latipes DNA, clone: olal-164P16.F, Genomic survey sequence.
ACCESSION  DE067970
VERSION     DE067970.1 GI:62579514
KEYWORDS   GSS.
SOURCE     Oryzias latipes (Japanese medaka)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianiichthyidae; Oryziinae; Oryzias.
REFERENCE  1

AUTHORS      Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.
TITLE        BAC end sequences of Olal Oryzias latipes Library
JOURNAL      Published Only in Database (2005)
REFERENCE    2 (bases 1 to 637)
AUTHORS      Fujiyama, A.
TITLE        Direct Submission
JOURNAL      Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
              (E-mail: afujiyam@gsc.riken.jp, URL: http://stt.gsc.riken.jp/,
              Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)
              This work was done in collaboration with Takeda, H. (1), Naruse, K.
              (2)
              and Narita, T. (3)
              (1) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: htakeda.s.u-tokyo.ac.jp
              (2) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: naruse.s.u-tokyo.ac.jp
              (3) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: tanarita.s.u-tokyo.ac.jp
              PRIMERS
              Sequencing : Forward
LIBRARY      Vector      : pKS145
              R.Site 1   : SacI
              L.Site 2   : SacI
              Location/Qualifiers
FEATURES     source
              1..637
              /organism="Oryzias latipes"
              /mol_type="genomic DNA"
              /db_xref="taxon:8090"
              /clone="olal-164P16.F"
              /sex="male"
              /cell_type="whole body"
              /clone_lib="BAC end sequences of Olal Oryzias latipes
              library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 11; Length 637;
Best Local Similarity 90.5%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTCATGGTTATCTTCC 21
    |||||
Db 518 TGCATGTCGCGTTTCTTCC 498

RESULT 8
LOCUS      BM645185/c
DEFINITION Oryzias latipes DNA, clone: olal-164P16.F, Genomic survey sequence.
ACCESSION  BM645185
VERSION     BM645185.1 GI:18944696
KEYWORDS   EST.
SOURCE     Anopheles gambiae (African malaria mosquito)
ORGANISM   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
REFERENCE  1 (bases 1 to 714)

```







```

QY      1  TGCATGTATGGGTTATCTT 19
      ||||| ||||| ||||| |||||
Db      43  TGCATATATGGGTTATCTT 25

RESULT 16
AW489071/c
LOCUS   397 bp      mRNA      linear      EST 24-FEB-2000
DEFINITION  UI-M-BH3-asd-h-03-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
              UI-M-BH3-asd-h-03-0-UI 3', mRNA sequence.
ACCESSION AW489071
VERSION   AW489071.1 GI:7059341
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
TITLE     1 (bases 1 to 397)
JOURNAL   Bonaldo,M.F., Lennon,G. and Soares,M.B.
PUBMED    Normalization and subtraction: two approaches to facilitate gene
8889548    discovery
COMMENT   Genome Res. 6 (9), 791-806 (1996)

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
1..397
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="UI-M-BH3-asd-h-03-0-UI"
    /dev_stage="27-32 days"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="NIH_BMAP_M_S4"
    /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site 1: Not I; Site 2: Eco RI; The
    NIH_BMAP_M_S4 library is a subtracted library of a series,
    ultimately derived from a mixture of individually tagged
    normalized libraries from ten regions of the mouse brain
    (cerebellum, brain stems, olfactory bulbs, hypothalamus,
    cortex, amygdala, basal ganglia, pineal gland, striatum,
    hippocampus) after a series of subtractions to reduce the
    representation of cDNAs from which ESTs had already been
    generated. The following serially subtracted libraries
    were generated in this process: NIH_BMAP_M_S4,
    NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
    NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
    (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
    cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
    NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
    was used as a driver in a hybridization with a pool of
    libraries in the form of single-stranded circles. The
    remaining single-stranded circles (subtracted library)
    was purified by hydroxyapatite column chromatography,
    converted to double-stranded circles and electroporated
    into DH10B bacteria (LifeTechnologies) to generate the

```

---

```

ORIGIN
Query Match      82.9%; Score 17.4; DB 1; Length 397;
Best Local Similarity 94.7%; Pred. No. 9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  GCATGTATGGGTTATCTTC 20
      ||||| ||||| ||||| |||||
Db      91  GCATGTATGGGTTATGTTTC 73

RESULT 17
AK219962
LOCUS    403 bp      mRNA      linear      HTC 23-NOV-2004
DEFINITION  Mus musculus cDNA, clone: Y2G0150E15, strand: minus,
              reference: ENSEMBL: Mouse-Transcript-ENST: ENSMUST00000061642, based
              on BLAT search.
ACCESSION  AK219962.1 GI:56044353
VERSION    AK219962
KEYWORDS   HTC; ASSETS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

REFERENCE
AUTHORS    Watahiki, A., Waki, K., Hayatsu, N., Shiraki, T., Kondo, S.,
           Nakamura, M., Sasaki, D., Arakawa, T., Kawai, J., Harbers, M.,
           Hayashizaki, Y. and Carninci, P.
TITLE      Libraries enriched for alternatively spliced exons reveal splicing
           patterns in melanocytes and melanomas
JOURNAL    Nat. Methods 1, 233-239 (2004)
REFERENCE  2 (bases 1 to 403)
AUTHORS    Arakawa, T., Carninci, P., Fukuda, S., Harbers, M., Hayatsu, N.,
           Hori, F., Imotani, K., Kawai, J., Kondo, S., Murata, M., Nakamura, M.,
           Nomura, K., Ohno, M., Sasaki, D., Shiraki, T., Waki, K., Watahiki, A. and
           Hayashizaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (15-SEP-2004) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
           URL: http://genome.gsc.riken.jp, Tel: 81-45-503-9222,
           Fax: 81-45-503-9216)
COMMENT    Alternative Splicing Libraries (ASLs) are prepared by: Preparing
           of single-stranded DNA using a RNA template from full length cDNA
           libraries, hybridizing of
           single-stranded DNAs, removing of remaining single-stranded DNA,
           bp-cutters, capturing of DNA hybrids with loop structures
           (alternative spliced exon), ligating of Y-shaped primers to
           isolated DNA hybrids with loop structures, PCR amplification of
           ligation products and their cloning into pFLCI vector. (Reference).
           Location/Qualifiers
           1..403
               /organism="Mus musculus"
               /mol_type="mRNA"
               /db_xref="taxon:10090"
               /clone="Y2G0150E15"
               /cell_lines="mixture of B16-F10Y and melan-c"
               /cell_type="mixture of melanoma cell and melanocyte cell"
               /clone_libs="Alternative Splicing Library L3"
               /note="strand: minus, reference: ENSEMBL: Mouse-Transcript-
               ENST: ENSMUST00000061642, based on BLAT search"
ORIGIN

```

NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_TISSUE=hypothalamus

TAG\_LIB=NIH\_BMAP\_M\_S4

TAG\_SEQ=CGGTA"

Query Match 82.9%; Score 17.4; DB 4; Length 403;  
Best Local Similarity 94.7%; Pred. No. 9.e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
|||||  
Db 366 GCATGTATGGTTATGTTTC 384

RESULT 18  
CB799737  
LOCUS CB799737 416 bp mRNA linear EST 16-MAY-2003  
DEFINITION AMGNNUC:NRWA3-00059-H10-A white adipose tiss (10469) Rattus  
norvegicus cDNA clone nrwa3-00059-h10 5', mRNA sequence.

ACCESSION CB799737  
VERSION CB799737.1 GI:29911338  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 416)  
AUTHORS Angen EST Program.  
TITLE Angen Rat EST Program  
JOURNAL Unpublished (2003)  
COMMENT Contact: Dan Fitzpatrick  
Angen, Inc

One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA  
Tel: 805 447-4881  
Plate: 00059 row: h column: 10.

Location/Qualifiers

## FEATURES

source

1..416  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="nrwa3-00059-h10"  
/tissue\_type="adipose tiss"  
/clone\_lib="white adipose tiss (10469)"  
/note="vector: pSPOR1; Site\_1: SalI; Site\_2: NotI; white  
adipose tiss adult female Wistar rats, >250 grams"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 416;  
Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
|||||  
Db 151 GCATGTATGGTTATGTTTC 169

RESULT 19  
AW492409/c  
LOCUS AW492409 429 bp mRNA linear EST 24-FEB-2000  
DEFINITION UI-M-BH3-atu-c-05-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-atu-c-05-0-UI 3', mRNA sequence.

ACCESSION AW492409  
VERSION AW492409.1 GI:7062690  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 429)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
PUBMED 8889548

COMMENT Contact: Chin, H  
National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov

Oligo-dt track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

## Location/Qualifiers

1..429  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-atu-c-05-0-UI"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_TISSUE=pineal-glands  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_SEQ=CAGAC"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 1; Length 429;  
Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
|||||  
Db 91 GCATGTATGGTTATGTTTC 73

RESULT 20  
CB738293

LOCUS CB738293

DEFINITION CB738293

ACCESSION CB738293

VERSION CB738293.1

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

CB738293 459 bp mRNA linear EST 11-APR-2003  
AMGNNUC:NRHY5-00423-B4-A W Rat hypothalamus (10471) Rattus  
norvegicus cDNA clone nrhy5-00423-b4 5', mRNA sequence.

CB738293  
CB738293.1 GI:29805558  
EST.  
Rattus norvegicus (Norway rat)

```
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS      Angen EST Program
TITLE        Angen Rat EST Program
JOURNAL      Unpublished (2003)
COMMENT      Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00423 row: b column: 4.

FEATURES
source       Location/Qualifiers
1. .459
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrhvs-00423-b4"
/clone_lib="W Rat hypothalamus (10471)"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; W Rat
hypothalamus adult female Wistar rat avg. insert size 2.3
kb fraction 6 and 7"

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 459;
Best Local Similarity 94.7%; Pred. No. 9.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GCATGTATGGGTATCTTC 20
        |||||
Db      268 GCATGTATGGGTATCTTC 286
        |||||

RESULT 21
LOCUS      CA895102
DEFINITION 481 bp mRNA linear EST 20-DEC-2002
Library (Long) Mus musculus cDNA clone NIA:B0189H07 IMAGE:30102426
5', mRNA sequence.
ACCESSION  CA895102
VERSION    CA895102.1 GI:27346651
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 481)
AUTHORS    Piao,Y., Dudekula,D.B., Qian,Y., Martin,P.R., Aiba,K., Vescovi,A.L.
and Ko,M.S.H.
TITLE      Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)
JOURNAL    Unpublished (2002)
COMMENT    Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0189 row: H column: 07
Seq primer: -21M13 Reverse
High quality sequence stop: 481
POLYA=No.

FEATURES
source       Location/Qualifiers
1. .481
/mol_type="mRNA"
/strain="CD1"
/db_xref="niaEST:B0189H07-5N"
/db_xref="taxon:10090"
/clone="NIA:B0189H07 IMAGE:30102426"
/dev_stage="Adult"
/lab_host="DH10B"

/clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Angelo L. Vescovi (Institute for Stem
Cell Research, Italy). Double-stranded cDNAs were:
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACTAGTTCGACGAGCGCGCCCTTTTTC-3'] from
2.0 Microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
3.2 kb. The library was constructed by Yulan Piao."
```



```

/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Angelo L. Vescovi (Institute for Stem
Cell Research, Italy). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACGTAGTCTAGATCGGAGCGCGCCCTTTTCTTTT-3'] from
2.0 Microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
3.2 kb. The library was constructed by Yulan Piao."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 553;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTC 20
      |||||
Db      277 GCATGTATGGTTATGTTC 295

RESULT 23
CA637985      561 bp      mRNA      linear      EST 23-NOV-2002
LOCUS      wreln.pk0005.h2 wreln Triticum aestivum cDNA clone wreln.pk0005.h2
DEFINITION      5' end, mRNA sequence.
ACCESSION      CA637985
VERSION      CA637985.1 GI:25216281
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 561)
AUTHORS      Tingey,S.V., Powell,W., Walters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      Dupont Wheat cDNA Sequence
JOURNAL      Unpublished (2002)
COMMENT      Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES
source
Location/Qualifiers
1..561
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wreln.pk0005.h2"
/tissue_type="root"
/clone_lib="wreln"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:

/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Angelo L. Vescovi (Institute for Stem
Cell Research, Italy). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACGTAGTCTAGATCGGAGCGCGCCCTTTTCTTTT-3'] from
2.0 Microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
3.2 kb. The library was constructed by Yulan Piao."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 553;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTC 20
      |||||
Db      277 GCATGTATGGTTATGTTC 295

RESULT 23
CA637985      561 bp      mRNA      linear      EST 23-NOV-2002
LOCUS      wreln.pk0005.h2 wreln Triticum aestivum cDNA clone wreln.pk0005.h2
DEFINITION      5' end, mRNA sequence.
ACCESSION      CA637985
VERSION      CA637985.1 GI:25216281
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 561)
AUTHORS      Tingey,S.V., Powell,W., Walters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      Dupont Wheat cDNA Sequence
JOURNAL      Unpublished (2002)
COMMENT      Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES
source
Location/Qualifiers
1..561
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wreln.pk0005.h2"
/tissue_type="root"
/clone_lib="wreln"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:

```

```

XhoI; Wheat (Triticum aestivum L.) root; normalized from
wre1 library"

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 561;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CATGTATGGTTATCTTCC 21
      |||||
Db      301 CATGTATGGTTATCTACC 319

RESULT 24
BQ571868      581 bp      mRNA      linear      EST 19-JUN-2002
LOCUS      UI-M-FCO-byc-b-20-0-UI.r1 NIH_BMAP_FCO Mus musculus cDNA clone
DEFINITION      IMAGE:5716147 5', mRNA sequence.
ACCESSION      BQ571868
VERSION      BQ571868.1 GI:21475185
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 581)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.

FEATURES
source
Location/Qualifiers
1..581
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6"
/clone="IMAGE:5716147"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FCO"
/notes="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA
tail, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 5; Length 581;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 GCATGTATGGGTATCTTC 20
Db      226 GCATGTATGGGTATCTTC 244

RESULT 25
BQ571429
LOCUS   589 bp mRNA linear EST 19-JUN-2002
DEFINITION
| | | | | | | | | | | | | | | |
IMAGE: 5715862 5', mRNA sequence.

ACCESSION BQ571429
VERSION   BQ571429.1 GI:21474746
KEYWORDS  ESI.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL   Sciurognathi; Muroidea; Muridae; Murinae; Mus.
COMMENT   1 (bases 1 to 589)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapbs@email.nih.gov
Tissue procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source
     1..589
         /organism="Mus musculus"
         /mol_type="mRNA"
         /strain="C57BL/6"
         /db_xref="taxon:10090"
         /clone="IMAGE:5715862"
         /tissue_type="whole brain"
         /dev_stage="embryo 12.5 dpc"
         /lab_host="DH10B (T1 phage resistant)"
         /clone_lib="NIH_BMAP_FCO"
         /notes="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;
         Site 2: Not I; The library was constructed according to
         Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
         1996. Denatured mRNA was size fractionated on a 1% agarose
         gel. First strand cDNA synthesis was primed with an
         oligo-dT primer containing a Not I site. Double stranded
         cDNA was size selected according to mRNA size fraction,
         ligated with Ecor I adaptor, digested with Not I, and then
         cloned directionally into pYX-Asc vector. The library tag
         sequence located between the Not I site and the polyA
         tail, is TGAAGAGACC. This library was created for the
         University of Iowa Mouse Brain Molecular Anatomy Project
         (BMAP). 'Gene Discovery in the Developing Mouse Nervous
         System', supported by National Institutes of Mental Health
         (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 5; Length 589;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 20
Db      331 GCATGTATGGGTATCTTC 349

RESULT 26

```

```

CE495411/c
LOCUS   592 bp DNA linear GSS 28-SEP-2003
DEFINITION
| | | | | | | | | | | | | | | |
IMAGE: 17000310260727 Dog Library Canis familiaris genomic,
genomic survey sequence.

ACCESSION CE495411
VERSION   CE495411.1 GI:36812192
KEYWORDS  GSS.
SOURCE    Canis familiaris (dog)
ORGANISM  Canis familiaris

REFERENCE
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
TITLE     Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Frazer,C.M. and
JOURNAL   Venter,J.C.
PUBMED    14512627
COMMENT   The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
     1..592
         /organism="Canis familiaris"
         /mol_type="genomic DNA"
         /strain="Standard Poodle"
         /db_xref="taxon:9615"
         /clone_lib="Dog Library"
         /note="Site 1: BstXI; Libraries were prepared from
         peripheral blood"

ORIGIN
Query Match      82.9%; Score 17.4; DB 10; Length 592;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTT 19
Db      94 TGCATGTATGGGTATGTT 76

RESULT 27
BZ300676/c
LOCUS   594 bp DNA linear GSS 31-OCT-2002
DEFINITION
| | | | | | | | | | | | | | | |
IMAGE: BZ300676 Kluyveromyces delphensis genomic clone KD0856, genomic survey
sequence.

ACCESSION BZ300676
VERSION   BZ300676.1 GI:24445706
KEYWORDS  GSS.
SOURCE    Kluyveromyces delphensis
ORGANISM  Kluyveromyces delphensis

REFERENCE
AUTHORS   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
TITLE     Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
JOURNAL   1 (bases 1 to 594)
PUBMED    12620120
COMMENT   Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie

QY      1 TGCATGTATGGGTATCTT 19
Db      94 TGCATGTATGGGTATGTT 76

RESULT 27
BZ300676/c
LOCUS   594 bp DNA linear GSS 31-OCT-2002
DEFINITION
| | | | | | | | | | | | | | | |
IMAGE: BZ300676 Kluyveromyces delphensis genomic clone KD0856, genomic survey
sequence.

ACCESSION BZ300676
VERSION   BZ300676.1 GI:24445706
KEYWORDS  GSS.
SOURCE    Kluyveromyces delphensis
ORGANISM  Kluyveromyces delphensis

REFERENCE
AUTHORS   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
TITLE     Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
JOURNAL   1 (bases 1 to 594)
PUBMED    12620120
COMMENT   Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie

```

```

FEATURES
  source
    Class: plasmid ends.
    Location/Qualifiers
      1..594
        /organism="Kluyveromyces delphensis"
        /mol_type="genomic DNA"
        /strain="CBS 2170"
        /db_xref="taxon:51657"
        /clone="KD0856"
        /clone_lib="Kluyveromyces delphensis Random Genomic
        Library"

ORIGIN
  Query Match      82.9%; Score 17.4; DB 9; Length 594;
  Best Local Similarity 94.7%; Pred. No. 9.5e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTT 19
    ||||| ||||| |||||
Db 411 TGCATGATGGGTATCTT 393

RESULT 28
CE104077      596 bp DNA linear GSS 24-SEP-2003
LOCUS        tigr-gss-dog-17000371142608 Dog Library Canis familiaris genomic,
DEFINITION   genomic survey sequence.
ACCESSION   CE104077
VERSION     CE104077.1 GI:35170962
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1 (bases 1 to 596)
AUTHORS    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
PUBMED     14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
            Location/Qualifiers
              1..596
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Standard Poodle"
                /db_xref="taxon:9615"
                /clone_lib="Dog Library"
                /note="Site 1: BstXI; Libraries were prepared from
                peripheral blood"

ORIGIN
  Query Match      82.9%; Score 17.4; DB 9; Length 596;
  Best Local Similarity 94.7%; Pred. No. 9.5e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTT 19
    ||||| ||||| |||||
Db 191 TGCATGATGGGTATCTT 209

RESULT 29
BE336403      600 bp mRNA linear EST 14-JUL-2000
LOCUS        601086981F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3481653 5',
DEFINITION   genomic survey sequence.
ACCESSION   CE133541
VERSION     CE133541.1 GI:35235919
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1 (bases 1 to 643)
AUTHORS    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
PUBMED     14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research

```

# FEATURES source

```

mRNA sequence.
BE336403      643 bp DNA linear GSS 25-SEP-2003
LOCUS        tigr-gss-dog-17000326280921 Dog Library Canis familiaris genomic,
DEFINITION   genomic survey sequence.
ACCESSION   CE133541
VERSION     CE133541.1 GI:35235919
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1 (bases 1 to 643)
AUTHORS    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
PUBMED     14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research

```

## FEATURES source

```

1..600
/organism="Mus musculus"
/mol_type="mRNA"
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3481653"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN
  Query Match      82.9%; Score 17.4; DB 2; Length 600;
  Best Local Similarity 94.7%; Pred. No. 9.5e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGGTATCTTC 20
    ||||| ||||| |||||
Db 307 GCATGTATGGGTATCTTC 325

```

## RESULT 30

```

CE133541
LOCUS        tigr-gss-dog-17000326280921 Dog Library Canis familiaris genomic,
DEFINITION   genomic survey sequence.
ACCESSION   CE133541
VERSION     CE133541.1 GI:35235919
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1 (bases 1 to 643)
AUTHORS    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
PUBMED     14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research

```

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

# Location/Qualifiers

```
1. .643
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: SetX1; Libraries were prepared from
peripheral blood"
```

# FEATURES

source

# ORIGIN

Query Match 82.9%; Score 17.4; DB 9; Length 643;  
Best Local Similarity 94.7%; Pred. No. 9.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTT 19  
|||||  
Db 570 TGCATGTATGGGTATGTT 588

RESULT 31  
BE533917  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE533917 685 bp mRNA linear EST 09-AUG-2000  
601234117F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3597604 5',  
mRNA sequence.  
BE533917  
BE533917.1 GI:9762562  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM8777 row: b column: 05  
High quality sequence stop: 586.

# FEATURES

source

```
1. .685
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam6"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"
```

# ORIGIN

Query Match 82.9%; Score 17.4; DB 2; Length 685;

Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 20  
|||||  
Db 17 GCATGTATGGGTATGTT 35

# RESULT 32

CA318453  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

UI-M-FW0-cbs-i-10-0-UI-r1 NIH BMAP\_FW0 Mus musculus cDNA clone  
IMAGE:6813875 5', mRNA sequence.  
CA318453  
CA318453.1 GI:24536577  
EST.  
Mus musculus (house mouse)

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

# REFERENCE

1 (bases 1 to 691)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

# Seq primer: pYX-5.

# Location/Qualifiers

source

```
1. .891
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_lib="NIH BMAP_FW0"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FW0"
/notes="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction. ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Aac vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AGCAGACAG. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."
```

# ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 691;

Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 20  
|||||  
Db 56 GCATGTATGGGTATGTT 74

```

RESULT 33
CF749756          697 bp  mRNA      linear      EST 10-OCT-2003
LOCUS             UI-M-HJO-cmn-o-24-0-UI.r1 NIH BMAP_HJO Mus musculus cDNA clone
DEFINITION        IMAGE:30628103 5', mRNA sequence.

ACCESSION         CF749756
VERSION           CF749756.1 GI:37646101
KEYWORDS          EST.
SOURCE            Mus musculus (house mouse)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                  Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-r@mail.nih.gov
                  Tissue Procurement: Dr. James Lin University of Iowa
                  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                  Clone Distribution: Distribution information can be found at
                  http://genome.uiowa.edu/distribution/mousefl.html
                  This clone was contributed by the Brain Molecular Anatomy Project
                  (BMAP)
FEATURES
  source          Location/Qualifiers
    1..697
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:30628103"
      /tissue_type="Upper Head"
      /dev_stage="9.5 and 10.5 dpc"
      /lab_host="DH10B (T1 phage resistant)"
      /clone_lib="NIH BMAP HJO"
      /note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;
        Site 2: Not I; The library was constructed according
        Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
        1996. Denatured RNA was size fractionated on a 1% agarose
        gel. First strand cDNA synthesis was primed with oligo-dT
        primer containing a Not I site. Double strand cDNA was
        size selected according to mRNA size fraction, ligated
        with EcoR I adaptor, digested with NotI and then cloned
        directionally into pYX-Asc vector. The library tag
        sequence located between the Not I site and the polyA tail
        is CGAACTGAAT. This library was created for the University
        Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
        Developing Mouse Nervous System', supported by National
        Institute of Mental Health (NIMH)."
```

```

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 697;
Best Local Similarity 94.7%; Pred. No. 9.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2  GCATGTATGGGTATCTTC 20
      |||||
Db   57 GCATGTATGGGTATGTTC 75
```

```

RESULT 34
CA511105          702 bp  mRNA      linear      EST 15-NOV-2002
LOCUS             UI-R-FJ0-cpv-e-20-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
DEFINITION        UI-R-FJ0-cpv-e-20-0-UI 5', mRNA sequence.
ACCESSION         CA511105
VERSION           CA511105.1 GI:25002059
KEYWORDS          EST.
```

## SOURCE ORGANISM

## REFERENCE AUTHORS TITLE

## JOURNAL PUBMED COMMENT

## FEATURES source

```

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 702)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 REVERSE.
Location/Qualifiers
  1..702
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="UI-R-FJ0-cpv-e-20-0-UI"
    /tissue_type="embryo"
    /dev_stage="embryo"
    /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-R-FJ0"
    /note="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;
      UI-R-FJ0 is a cDNA library containing the following
      tissue(s): rat embryo. The library was constructed
      according to Bonaldo, Lennon and Soares, Genome Research,
      6:791-806, 1996. First strand cDNA synthesis was primed
      with an oligo-dT primer containing a Not I site. Double
      stranded cDNA was ligated to an EcoR I adaptor, digested
      with Not I, and cloned directionally into pYX-Asc
      vector. The oligonucleotide used to prime the synthesis of
      first-strand cDNA contains a library tag sequence that is
      located between the Not I site and the dT18 tail. The
      sequence tag for this library is CATCTCTACT. This library
      was created for the University of Iowa Program for Rat
      Gene Discovery and Mapping (Val Sheffield, Bento Soares
      and Tom Casavant)"
```

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 6; Length 702;
Best Local Similarity 94.7%; Pred. No. 9.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2  GCATGTATGGGTATCTTC 20
      |||||
Db   129 GCATGTATGGGTATGTTC 147
```

```

RESULT 35
CB519329          707 bp  mRNA      linear      EST 09-JUL-2003
LOCUS             UI-M-GH0-cef-e-10-0-UI.r1 NIH BMAP_GH0 Mus musculus cDNA clone
DEFINITION        IMAGE:6838739 5', mRNA sequence.
ACCESSION         CB519329
VERSION           CB519329.1 GI:29352684
KEYWORDS          EST.
SOURCE            Mus musculus (house mouse)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES  
 source

Location/Qualifiers

1..707  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6838739"  
 /tissue\_type="whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP GH0"  
 /notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaïdo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGCTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 707;  
 Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTATCTTC 20

|||||  
 Db 170 GCATGTATGGTATCTTC 188

RESULT 36

CV805733/c

LOCUS

DEFINITION CV805733 714 bp mRNA linear EST 21-JAN-2005  
 IMAGE:7465640 5', mRNA sequence.

ACCESSION

CV805733

VERSION

CV805733.1

KEYWORDS

EST.

SOURCE

Xenopus tropicalis (western clawed frog)

ORGANISM

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus; Silurana.

REFERENCE

1 (bases 1 to 714)

NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics

National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Bruce Blumberg  
 cDNA Library preparation: B. Blumberg  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM15760 row: b column: 06  
 High quality sequence start: 6  
 High quality sequence stop: 282.

FEATURES

source

Location/Qualifiers

1..714  
 /organism="Xenopus tropicalis"  
 /mol\_type="mRNA"  
 /strain="N6 (Nigerian 6th generation inbred)"  
 /db\_xref="taxon:8364"  
 /clone="IMAGE:7465640"  
 /tissue\_type="Kidney"  
 /dev\_stage="Adult"  
 /lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"  
 /clone\_lib="NIH XGC tropKid1"  
 /notes="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming (5'-GAGAGAGAGAGAGAGACTAGTCTCGAGTTTTTTTTTTT-3') and Stratascript reverse transcriptase. After ligation of EcoRI adapters (5'-AATTCGACAGG-3') followed by kinasing adapters and by XhoI digestion, the cDNA was size selected by chromatography on Sepharose CL-2B columns and fractions containing cDNAs larger than 1000 bp were ligated into EcoRI/XhoI-digested pCS107. Reference for library construction: Current Genomics 4, 635-644. Library constructed by Michelle Tabb and Bruce Blumberg (Dept of Developmental and Cell Biology, University of California, Irvine)."

ORIGIN

Query Match 82.9%; Score 17.4; DB 8; Length 714;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTATCTT 19

|||||  
 Db 359 TGCATGTATGTATCTT 341

RESULT 37

CK364526

LOCUS

DEFINITION CK364526 721 bp mRNA linear EST 23-DEC-2003  
 IMAGE:7095948 5', mRNA sequence.

ACCESSION

CK364526

VERSION

CK364526.1

KEYWORDS

EST.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 721)

NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM14942 row: h column: 02  
 High quality sequence stop: 722.  
 Location/Qualifiers

#### FEATURES

1..721  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="IMAGE:7096948"  
 /tissue\_type="lung, pooled"  
 /lab\_host="DH10B Tora"  
 /clone\_lib="NIH\_MGC\_231"  
 /notes="Organ: lung; Vector: pExpress-1; Site: 1: EcoRV;  
 Site 2: NotI; RNA obtained from pooled lung tissue from a  
 mix of male and female animals at 8 wk old. Tissues were  
 snap-frozen and kept at -80C for two days before RNA  
 extraction and purification (Tri-reagent method). cDNA was  
 primed using oligo-dT primer:  
 5'-pGATGTTCTAGTCGAGCGCGCCCTT)25-3' and cloned into  
 the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb  
 resulted in an average insert size of 2.3 kb. This primary  
 library is not normalized (normalized primary library is  
 NIH MGC 232) and was constructed by Express Genomics  
 (Frederick, MD). Note: this is a NIH\_MGC library."

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 721;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTATCTTC 20  
 |||||  
 Db 636 GCATGTATGGTATGTC 654

RESULT 38  
 CB245585  
 LOCUS  
 DEFINITION UI-M-FYO-cdt-o-22-0-UI.r1 NIH\_BMAP\_FYO Mus musculus cDNA clone  
 IMAGE:6834383 5', mRNA sequence.

ACCESSION CB245585  
 VERSION CB245585.1 GI:28367229  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 767)  
 NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

#### REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

#### FEATURES

Seq primer: pYX-5.  
 Location/Qualifiers  
 1..767  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"

/db\_xref="taxon:10090"  
 /clone="IMAGE:6834383"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_FYO"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 to the protocol of Bontolillo et al. (1996). Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 is ACGGACACAG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the  
 Developing Mouse Nervous System, supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 767;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTATCTTC 20  
 |||||  
 Db 292 GCATGTATGGTATGTC 310

#### RESULT 39

CO428850  
 LOCUS  
 DEFINITION UI-M-HXO-crw-m-17-0-UI.r1 NIH\_BMAP\_HXO Mus musculus cDNA clone  
 IMAGE:30686416 5', mRNA sequence.

ACCESSION CO428850  
 VERSION CO428850.1 GI:49675144  
 KEYWORDS EST.

#### SOURCE

Mus musculus (house mouse)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 768)  
 NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

The following repetitive elements were found in this cDNA  
 sequence: 175-222, >(CAG)n#Simple\_repeat (matched complement)  
 Seq primer: pYX-5.  
 Location/Qualifiers  
 1..768  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"

#### FEATURES

source  
 1..768  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30686416"  
 /tissue\_type="whole eye"  
 /dev\_stage="newborn (1, 5, 15 days) and embryonic (15, 16,  
 17, 18 dpc)"  
 /lab\_host="DH10B (T1 phage resistant)"

/clone lib="NIH\_BMAP\_HX0"  
 /notes="Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I;  
 Site\_2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAAATTACG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 768;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 664 GCATGTATGGTTATCTTC 682

## RESULT 40

CD349045 777 bp mRNA linear EST 09-JUL-2003  
 LOCUS UI-M-FYO-cfp-g-14-0-UI.r1 NIH\_BMAP\_FYO Mus musculus cDNA clone  
 DEFINITION IMAGE:6854151 5', mRNA sequence.

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 664 GCATGTATGGTTATCTTC 682

CD349045.1 GI:31140560  
 Mus musculus (house mouse)

EST.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 777)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

1..777

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:6854151"

/tissue\_type="whole brain"

/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_FYO"

/notes="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;  
 Site 2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is ACGGAGACAG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 777;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 23 GCATGTATGGTTATCTTC 41

## RESULT 41

CX563730 787 bp mRNA linear EST 12-JAN-2005  
 LOCUS UI-M-IB0-cue-d-03-0-UI.r1 NIH\_BMAP\_IB0 Mus musculus cDNA clone  
 DEFINITION IMAGE:30940682 5', mRNA sequence.

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 23 GCATGTATGGTTATCTTC 41

CX563730.1 GI:57590759

EST.

ORGANISM

Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 787)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..787

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:30940682"

/tissue\_type="whole eye"

/dev\_stage="newborn (1,5,15 days)"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_IB0"

/notes="Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I;  
 Site\_2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAAATTACG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN



Query Match 82.9%; Score 17.4; DB 8; Length 787;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTCCTTC 20  
 |||||  
 Db 293 GCATGTATGGTTCCTTC 311

RESULT 42  
 CB521799  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 793)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
 source  
 Location/Qualifiers  
 1..793  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6842025"  
 /tissue\_type="whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_GHO"  
 /notes="Organ: Brain; Vector: pYX-Asc; Site\_1: Ecor I;  
 Site\_2: Not I; The library was constructed according  
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CGACTGAGAT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

Query Match 82.9%; Score 17.4; DB 6; Length 793;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTCCTTC 20  
 |||||

Db 571 GCATGTATGGTTCCTTC 589

RESULT 43  
 CF737017  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 797)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
 source  
 Location/Qualifiers  
 1..797  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30608409"  
 /tissue\_type="whole eye"  
 /dev\_stage="embryo 12.5, 13.5, 14.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_HDO"  
 /notes="Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I;  
 Site\_2: Not I; The library was constructed according  
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is TTATTGAAGT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), "

# ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 797;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 GCATGTATGGTTCCTTC 20  
 |||||  
 Db 181 GCATGTATGGTTCCTTC 199

RESULT 44  
 CB290090  
 LOCUS  
 DEFINITION  
 ACCESSION  
 CB290090  
 UI-M-F10-cdx-e-14-0-UI.r1 NIH BMAP\_F10 Mus musculus cDNA clone  
 IMAGE:6835671 5', mRNA sequence.

```

VERSION      CB290090.1  GI:28611505
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. James Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)

The following repetitive elements were found in this cDNA
sequence: 186-233, >(CAG)n$imple_repeat (matched complement)
Seq primer: pYX-5,

FEATURES
source       Location/Qualifiers
             1..802
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6"
             /db_xref="taxon:10090"
             /clone="IMAGE:6835671"
             /tissue_type="whole brain"
             /dev_stage="embryo 12.5dpc"
             /lab_host="DH10B (T1 phage resistant)"
             /clone_lib="NIH BMAP FTO"
             /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
             Site 2: Not I; The library was constructed according
             Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
             1996. Denatured RNA was size fractionated on a 1% agarose
             gel. First strand cDNA synthesis was primed with oligo-dT
             primer containing a Not I site. Double strand cDNA was
             size selected according to mRNA size fraction, ligated
             with EcoR I adaptor, digested with NotI and then cloned
             directionally into pYX-Asc vector. The library tag
             sequence located between the Not I site and the polyA tail
             is CAGCCACGAC. This library was created for the University
             Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
             Developing Mouse Nervous System', supported by National
             Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
             program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 802;
Best Local Similarity 94.7%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGATGGGTATCTTC 20
       |||||
Db      674 GCATGATGGGTATGTTTC 692

RESULT 45
LOCUS     BQ180689
DEFINITION
UI-M-EX0-bxc-c-05-0-UI_r1 NIH BMAP_EX0 Mus musculus cDNA clone
IMAGE:5706172 5', mRNA sequence.
BQ180689
ACCESSION BQ180689
VERSION    BQ180689.1  GI:20356181
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)

The following repetitive elements were found in this cDNA
sequence: 186-233, >(CAG)n$imple_repeat (matched complement)
Seq primer: pYX-5,

FEATURES
source       Location/Qualifiers
             1..804
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6"
             /db_xref="taxon:10090"
             /clone="IMAGE:5706172"
             /tissue_type="whole brain"
             /dev_stage="embryo 15.5 dpc"
             /lab_host="DH10B (T1 phage resistant)"
             /clone_lib="NIH BMAP EX0"
             /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
             Site 2: Not I; The library was constructed according to
             Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
             1996. Denatured mRNA was size fractionated on a 1% agarose
             gel. First strand cDNA synthesis was primed with an
             oligo-dT primer containing a Not I site. Double stranded
             cDNA was size selected according to mRNA size fraction,
             ligated with EcoR I adaptor, digested with Not I, and then
             cloned directionally into pYX-Asc vector. The library tag
             sequence located between the Not I site and the polyA
             tail, is GTGCGTGGA. This library was created for the
             University of Iowa Mouse Brain Molecular Anatomy Project
             (BMAP): 'Gene Discovery in the Developing Mouse Nervous
             System', supported by National Institutes of Mental Health
             (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 3; Length 804;
Best Local Similarity 94.7%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGATGGGTATCTTC 20
       |||||
Db      739 GCATGATGGGTATGTTTC 757

RESULT 46
LOCUS     CF726722
DEFINITION
UI-M-HB0-cki-i-01-0-UI_r1 NIH BMAP_HB0 Mus musculus cDNA clone
IMAGE:30548448 5', mRNA sequence.
CF726722
ACCESSION CF726722
VERSION    CF726722.1  GI:37600890
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
The following repetitive elements were found in this cDNA  
sequence: 96-143, >(CAG)n#simple\_repeat (matched complement)  
Seq primer: pYX-5.

FEATURES  
source

Location/Qualifiers  
1. .806  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30548448"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP HB0"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATGTAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 806;  
Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GCATGTATGGTTATCTTC 20  
|||||  
Db 584 GCATGTATGGTTATGTC 602

RESULT 47  
BQ444317  
LOCUS  
DEFINITION  
BQ444317  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BQ444317 814 bp mRNA linear EST 29-MAY-2002  
UI-M-EXO-bxk-k-22-0-UI-r1 NIH BMAP\_EXO Mus musculus cDNA clone  
IMAGE:5709453 5', mRNA sequence.

BQ444317.1 GI:21247429  
EST.  
Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 814)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

## FEATURES

source

Location/Qualifiers  
1. .814  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5709453"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP EX0"  
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GTGCGTGGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 5; Length 814;  
Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GCATGTATGGTTATCTTC 20  
|||||  
Db 46 GCATGTATGGTTATGTC 64

RESULT 48  
BI685030

LOCUS  
DEFINITION  
BI685030 840 bp mRNA linear EST 18-SEP-2001  
603310009F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5345930 5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BI685030  
BI685030.1 GI:15647658  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 840)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.  
cDNA Library preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM11879 row: a column: 03  
High quality sequence stop: 813.

FEATURES  
source

Location/Qualifiers  
1. .840

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5345930"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam6"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Salt; Site 2: Not I; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

```

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 3; Length 840;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 GCATGATGGGTATCTTC 20
      |||||
DB      739 GCATGATGGGTATCTTC 757

```

## RESULT 49

```

BU613524
LOCUS
DEFINITION
  BU613524 845 bp mRNA linear EST 20-FEB-2003
  UI-M-EMO-cas-p-10-0-UI.r1 NIH BMAP EMO Mus musculus cDNA clone
  UI-M-EMO-cas-p-10-0-UI 5', mRNA sequence.
ACCESSION
  BU613524
VERSION
  BU613524.1 GI:23279739
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

## REFERENCE

```

AUTHORS
  NIH-MGC http://mgc.nci.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: gcapbs-remail.nih.gov
  Tissue Procurement: Dr. James Lin, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bentso-soares@uiowa.edu
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)

```

Seq primer: pYX-5.

## FEATURES

## source

```

Location/Qualifiers
1..845
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="UI-M-EMO-cas-p-10-0-UI"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EMO"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag

```

sequence located between the Not I site and the polyA tail, is GTGCGTGA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 5; Length 845;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 GCATGATGGGTATCTTC 20
      |||||
DB      441 GCATGATGGGTATCTTC 459

```

## RESULT 50

```

CNS0460Q/c
LOCUS
DEFINITION
  CNS0460Q 961 bp DNA linear GSS 01-SEP-2000
  Tetraodon nigroviridis genome survey sequence T7 end of clone
  086K03 of library G from Tetraodon nigroviridis, genomic survey
  sequence.
ACCESSION
  AL277019,1 GI:8011222
VERSION
  GSS; genome survey sequence.
KEYWORDS
  Tetraodon nigroviridis
SOURCE
  Tetraodon nigroviridis
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
  Tetraodontidae; Tetraodontidae; Tetraodon.

```

## REFERENCE

```

AUTHORS
  Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
  Bernot,A., Fzames,C., Wincker,P., Brottier,P., Quetier,F.,
  Saurin,W. and Weissbach,J.
TITLE
  Estimate of human gene number provided by genome-wide analysis
  using Tetraodon nigroviridis DNA sequence
JOURNAL
  Nat. Genet. 25 (2), 235-238 (2000)
PUBMED
  10835645
REFERENCE
  2
  Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
  Fzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
  Saurin,W., Bernot,A. and Weissbach,J.
  Characterization and repeat analysis of the compact genome of the
  freshwater pufferfish Tetraodon nigroviridis
  Genome Res. 10 (7), 939-949 (2000)
PUBMED
  10899143
REFERENCE
  3 (bases 1 to 961)
AUTHORS
  Genoscope.

```

## TITLE

```

Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

```

## FEATURES

## source

```

Location/Qualifiers
1..961
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="086K03"
/clone_lib="G"
/notes="Genoscope sequence ID : COBG086AF02LP1
end : T7"

```

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 11; Length 961;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  TGCATGTATGGGTTATCTCC 21
      ||||| ||||| ||||| ||||| |||||
Db      623 TGCATGTCTGGGTTTCTYCC 603

RESULT 51
BG519143      995 bp      mRNA      linear      EST 02-APR-2001
DEFINITION   602578081F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3481653 5',
mRNA sequence.
ACCESSION   BG519143
VERSION     BG519143.1 GI:13514654
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 995)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-x@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM8510 row: b column: 22
            High quality sequence stop: 741.
            Location/Qualifiers
                1..995
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="mix FVB/N, C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:3481653"
                /tissue_type="tumor, gross tissue"
                /dev_stage="7 months"
                /lab_host="DH10B"
                /clone_lib="NCI_CGAP_Mam5"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Robin Humphreys,
                NIH"

ORIGIN
Query Match      82.9%; Score 17.4; DB 2; Length 995;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  GCATGTATGGGTTATCTTC 20
      ||||| ||||| ||||| ||||| |||||
Db      326 GCATGTATGGGTTATGTTTC 344

RESULT 52
CL116220      1032 bp      DNA      linear      GSS 05-JAN-2004
LOCUS        ISB1-63M20.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-63M20,
DEFINITION   genomic survey sequence.
ACCESSION   CL116220
VERSION     CL116220.1 GI:40609855
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
ORGANISM    Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.

REFERENCE   1  (bases 1 to 1032)
AUTHORS     Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE       A physical map of the xenopus tropicalis genome
JOURNAL     Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000 Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 8
            High quality sequence stop: 374.
            Location/Qualifiers
                1..1032
                /organism="Xenopus tropicalis"
                /mol_type="genomic DNA"
                /db_xref="taxon:8364"
                /clone="ISB1-63M20"
                /clone_lib="ISB1"
                /note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC
                Library Segment 1"

ORIGIN
Query Match      82.9%; Score 17.4; DB 10; Length 1032;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TGCATGTATGGGTTATCTTT 19
      ||||| ||||| ||||| ||||| |||||
Db      40  TGCATTATGGGTTATCTTT 58

RESULT 53
AK029006      3610 bp      mRNA      linear      HTC 03-APR-2004
LOCUS        Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DEFINITION   library, clone:4732480N13 product:DnaJ (Hsp40) homolog, subfamily
            C, member 5, full insert sequence.
ACCESSION   AK029006
VERSION     AK029006.1 GI:26324963
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  Carninci, P. and Hayashizaki, Y.
            High-efficiency full-length cDNA cloning
            Meth. Enzymol. 303, 19-44 (1999)
            10349636
REFERENCE   2  Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
            Genome Res. 10 (10), 1617-1630 (2000)
            11042159
REFERENCE   3  Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
            Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
            Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multipillar sequencer
            Genome Res. 10 (11), 1757-1771 (2000)
            11076861
REFERENCE   4

```

AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	RESULT 54	AK032373
TITLE	Functional annotation of a full-length mouse cDNA collection	LOCUS	AK032373
REFERENCE	Nature 409, 685-690 (2001)	DEFINITION	
AUTHORS	5		
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	ACCESSION	AK032373
JOURNAL	Analysis of the mouse transcriptome based on functional annotation	VERSION	AK032373.1
REFERENCE	Nature 420, 563-573 (2002)	KEYWORDS	HTC; CAP trapper.
AUTHORS	6 (bases 1 to 3610)	SOURCE	Mus musculus (house mouse)
		ORGANISM	Mus musculus
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Furuta, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	REFERENCE	1 Carninci, P. and Hayashizaki, Y. 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	TITLE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hara, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system - 384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
COMMENT	CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	JOURNAL	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
FEATURES	source	REFERENCE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
	1. .3610 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:4732480N13" /db_xref="taxon:10090" /clone="4732480N13" /tissue_type="skin" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days neonate" 163. .759 /notes="unnamed protein product; DnaJ (Hsp40) homolog, subfamily C, member 5 [MGD MGI:892995, GB NM_016775, evidence: BLASTN, 99%, match=597] putative" /protein_start=1 /protein_id="BAC26236.1" /db_xref="GI:26324964" /translations="MADQQRSLSTGESLYHLGLDKNATSDDIKSKYRKALKYHP DKPNDPEADKFEINNAHILTDATKNIYDKYSLGLYVAEQGEENVNTYFVL SWAKALFVVCGLTCCYCCCLCCFCNCCGCKPKAPGEETEFVSPEDLEAQLQ SDREATDTFIVQPASATTTQLTADSHFSYHTDGN"	JOURNAL	6 (bases 1 to 3626) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Furuta, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
	2 GCATGTATGGGTATCTTC 20       1001 GCATGTATGGGTATCTTC 1019	TITLE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)
ORIGIN	Query Match 82.9%; Score 17.4; DB 4; Length 3610; Best Local Similarity 94.7%; Pred. No. 1.2e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	COMMENT	RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers		TITLE	Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
FEATURES	source	REFERENCE	1. .3626
		AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
FEATURES	source	REFERENCE	5
		AUTHORS	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
FEATURES	source	REFERENCE	6
		AUTHORS	(bases 1 to 3732) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
FEATURES	source	COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers
		misc_feature	1. .3732 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM,DB:D030049H18" /db_xref="taxon:10090" /clone="D030049H18" /tissue_type="whole body" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="9 days embryo" 1. .3732 /notes="DnaJ (Hsp40) homolog, subfamily C, member 5 (MGD MGI:892995, GB NM_016775, evidence: BLASTN, 99%, match=597)"
FEATURES	source	ORIGIN	Query Match 82.9%; Score 17.4; DB 4; Length 3732; Best Local Similarity 94.7%; Pred.No.1.2e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  Qy 2 GCATGTATGGTTATCTTC 20       Db 995 GCATGTATGGTTATGTTTC 1013  RESULT 55 AK083584 LOCUS DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030049H18 product:DnaJ (Hsp40) homolog, subfamily C, member 5, full insert sequence. AK083584 VERSION AK083584.1 GI:26101348 KEYWORDS HTC; CAP trapper. SOURCE Mus musculus (house mouse) ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 10349636 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159 3 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

```

LOCUS
DEFINITION ZMMBB0321B11.f ZMMBB Zea mays genomic clone ZMMBB0321B11 5',
668 bp DNA linear GSS 22-OCT-2003
CG735819
genomic survey sequence.
ACCESSION CG735819
VERSION
KEYWORDS GSS.
SOURCE
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 668)
AUTHORS Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
and Wing, R.K.
TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: T7
BACKWARD: M13r
Plate: 0321 row: B column: 11
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..668
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:B73"
/clone="ZMMBB0321B11"
/lab_host="DH10B"
/clone_lib="ZMMBBb"
/notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII; Zea mays L. ssp. mays"
ORIGIN
Query Match 81.0%; Score 17; DB 10; Length 668;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTTATC 17
DB 174 TGCATGTATGGGTTATC 190
RESULT 57
CN212867/c
LOCUS CN212867 738 bp mRNA linear EST 08-APR-2004
DEFINITION 28396 Suspension culture Solanum tuberosum cDNA, mRNA sequence.
ACCESSION CN212867
VERSION
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE
1 (bases 1 to 738)
AUTHORS Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Laque, M., De
Koeyer, D., Audy, P., Goyer, C., Li, X.-Q., Wang-Pruski, G. and Regan, S.
Generation of ESTs from potato suspension cultures
JOURNAL Unpublished (2004)
COMMENT Contact: Barry Flinn
The Canadian Potato Genome Project - BioAtlantech
921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
Email: bflinn@bioatlantech.nb.ca
Seq primer: T3
Location/Qualifiers
1..738
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Shepody"
/db_xref="taxon:4113"
/tissue_type="Callus-derived suspension culture"
/clone_lib="X110-Gold"
/lab_host="X110-Gold"
/notes="Vector: pBluescript II SK(+) XR; Site_1: EcoRI;
Site_2: XhoI; supplier: Developmental series. Callus was
induced from Shepody, Clone 1756, sterile stem sections by
culture on Callus Induction Medium (CIM), comprised of MS
medium (pH 5.6) containing 10 mg/L Thiamine-HCL, .01 mg/L
Kinetin and 3 mg/L 2,4-D solidified with .8% (w/v)
Phytagar. Suspensions were induced by placing callus from
the plates into 125 ml Erlenmeyer flasks with liquid CIM
(no Phytagar) at a density of 10% (w/v) in volumes of
approximately 30-35 ml. Cells were subcultured weekly by
transfer to fresh media, with the density remaining at 10%
(w/v) and the volume remaining around 30 ml. Cells were
collected for RNA isolations and library construction 5
days after subculture."
ORIGIN
Query Match 81.0%; Score 17; DB 7; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 ATGTATGGGTATCTTC 20
DB 253 ATGTATGGGTATCTTC 237
RESULT 58
BW810341
LOCUS BW810341 109 bp mRNA linear EST 23-MAY-2005
DEFINITION BW810341 Amphioxus Branchiostoma floridae unpublished cDNA library,
larva whole animal Branchiostoma floridae cDNA clone b51v047c21 5',
mRNA sequence.
ACCESSION BW810341
VERSION BW810341.1 GI:66418557
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
REFERENCE
1 (bases 1 to 109)
AUTHORS Yu, J., Holland, L.Z., Shin-i, T., Kohara, Y., Satou, Y. and Satoh, N.
Expressed genes in Branchiostoma floridae
JOURNAL Unpublished (2005)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
Location/Qualifiers
1..109
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="b51v047c21"
/tissue_type="whole animal"
/dev_stage="larva"
/clone_lib="Amphioxus Branchiostoma floridae unpublished
cDNA library, larva whole animal"
ORIGIN
Query Match 80.0%; Score 16.8; DB 5; Length 109;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;

```



Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
 ||||| |  
 Db 52 TGCATGTATGGTTATCTTC 71

RESULT 59  
 BB303972 226 bp mRNA linear EST 10-JUL-2000  
 LOCUS BB303972 RIKEN full-length enriched, adult male corpora  
 DEFINITION quadrigenina Mus musculus cDNA clone B230107M10 3', mRNA sequence.  
 BB303972  
 ACCESSION BB303972  
 VERSION BB303972.1 GI:9004677  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 226)  
 KONNO, H., AIZAWA, K., AKAHIRA, S., AKIYAMA, J., ARAKAWA, T.,  
 CARNINCI, P., ENDO, T., FUKUDA, S., FUKUNISHI, Y., HARA, A., HAYATSU, N.,  
 HIRAZANE, T., HORI, F., ISHII, Y., ISHIKAWA, J., ISHIKAWA, T., ITOH, M.,  
 IZAWA, M., KADOTA, K., KAGAWA, I., KAI, C., KAWAI, J., KIKUCHI, N.,  
 KIYOSAWA, H., KOJIMA, Y., KONDO, S., KOYA, S., KURIHARA, C.,  
 KUSAKABE, M., MATSUYAMA, T., MIKI, R., MIZUNO, Y., NAKAMURA, M., ODA, H.,  
 OKAZAKI, Y., ONO, T., OWA, C., SAITO, H., SAKAI, C., SATO, K.,  
 SHIBATA, K., SHIBATA, Y., SHIGEMOTO, Y., SHINAGAWA, A., SHIRAKI, T.,  
 SUGABE, Y., SUGAHARA, Y., SUZUKI, H., SUZUKI, H., TAGAWA, A.,  
 TAKAHASHI, F., TOMINAGA, N., TOYA, T., TSUNODA, Y., WATAHIKI, A.,  
 WATANABE, S., YAMANAKA, T., YAMANAKA, I., YANO, R., YASUNISHI, A.,  
 YOKOTA, T., YOSHIDA, K., YOSHIKI, A., YOSHINO, M., MURAMATSU, M. and  
 HAYASHIZAKI, Y.  
 RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S.,  
 Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Thermotabilization and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
 Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (http://genome.rtc.riken.go.jp) for  
 further details.

FEATURES  
 source  
 1..226  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="B230107M10"  
 /sex="male"  
 /tissue\_type="corpora quadrigenina"  
 /dev\_stages="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, adult male corpora  
 quadrigenina"  
 /note="site 1: Sall; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia

## ORIGIN

Query Match 80.0%; Score 16.8; DB 2; Length 226;  
 Best Local Similarity 90.0%; Pred. No. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
 ||||| |

Db 14 TGCATGTATGGTTATCTTC 33  
 ||||| |

## RESULT 60

LOCUS CN900771/c 243 bp mRNA linear EST 04-JUN-2004  
 DEFINITION 010717ABBA001805HT (ABBA) Pacific Rose spur buds from OFF trees  
 Malus x domestica cDNA clone ABBA001805, mRNA sequence.  
 CN900771

ACCESSION CN900771.1 GI:48287012

VERSION EST.

KEYWORDS Malus x domestica

SOURCE Malus x domestica

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Eukaryota; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.

REFERENCE 1 (bases 1 to 243)

AUTHORS Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B.,  
 McArtney, S., Newcomb, R., Ross, G., Snowden, K., Walton, E. and Yauk, Y.

TITLE HortResearch Apple EST Project

JOURNAL Unpublished (2004)

COMMENT Contact: Gleave, A.  
 Sequencing Facility  
 The Horticulture and Food Research Institute of New Zealand Ltd  
 120 Mt Albert Rd, Mt Albert, Auckland, New Zealand  
 Tel: 00 64 09 815 4200  
 Fax: 00 64 09 815 4201  
 Email: est@hortresearch.co.nz.

## FEATURES

source

1..243  
 /organism="Malus x domestica"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3750"  
 /clone="ABBA001805"  
 /tissue\_type="Spur buds from 'OFF' trees"  
 /clone\_lib="(ABBA) Pacific Rose spur buds from OFF trees"  
 /note="vector: pBluescript SK(-); Library sequenced by  
 Genesis Research & Development"

## ORIGIN

Query Match 80.0%; Score 16.8; DB 7; Length 243;  
 Best Local Similarity 90.0%; Pred. No. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 21  
 ||||| |

Db 36 GGAATGTATGGTTATCTTC 17  
 ||||| |

Search completed: January 28, 2006, 01:14:17

Job time : 1712.94 secs

*This Page Blank (uspto)*

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 58,988 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattatgggttatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*

7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*

8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*

9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	3	US-10-081-923-2
2	21	100.0	1323	3	US-10-081-923-12
3	21	100.0	1378	3	US-10-081-923-14
4	21	100.0	1379	3	US-10-081-923-7
5	21	100.0	1379	3	US-10-081-923-13
6	21	100.0	1379	3	US-10-081-923-16
7	21	100.0	1384	3	US-10-081-923-8
8	21	100.0	1384	3	US-10-081-923-10
9	21	100.0	1385	3	US-10-081-923-9
10	21	100.0	1390	3	US-10-081-923-11
11	21	100.0	1393	3	US-10-081-923-15
12	21	100.0	1447	3	US-10-081-923-17
13	21	100.0	1803	3	US-10-081-923-5
14	19.4	92.4	1140	3	US-09-583-110-1191
15	19.4	92.4	1287	3	US-09-107-433-1469
16	19.4	92.4	8195	3	US-08-961-527-94
17	16.8	80.0	89210	3	US-09-949-016-16988
18	16.8	80.0	144922	3	US-09-949-016-15890
19	16.2	77.1	1059	3	US-09-328-352-540
20	16	76.2	1356	3	US-09-248-796A-2116
21	16	76.2	4291	3	US-09-351-200-1
22	15.8	75.2	601	3	US-09-949-016-24376
23	15.8	75.2	601	3	US-09-949-016-24696
24	15.8	75.2	601	3	US-09-949-016-111355

```

25 15.8 75.2 601 3 US-09-949-016-160998 Sequence 160998,
c 26 15.8 75.2 601 3 US-09-949-016-199547 Sequence 199547,
c 27 15.8 75.2 8351 2 US-08-198-446B-14 Sequence 14, Appl
c 28 15.8 75.2 8351 2 US-08-870-693-14 Sequence 14, Appl
c 29 15.8 75.2 9721 3 US-09-949-016-14814 Sequence 14814, A
c 30 15.8 75.2 9721 3 US-09-949-016-12076 Sequence 12076, A
c 31 15.8 75.2 26831 3 US-09-949-016-16250 Sequence 16250, A
c 32 15.8 75.2 36755 3 US-09-949-016-16994 Sequence 16994, A
c 33 15.8 75.2 38059 3 US-09-328-925-4 Sequence 4, Appl
c 34 15.8 75.2 41435 3 US-09-949-016-15926 Sequence 15926, A
c 35 15.8 75.2 152132 3 US-09-949-016-13845 Sequence 13845, A
c 36 15.8 75.2 152145 3 US-09-949-016-12371 Sequence 12371, A
c 37 15.8 75.2 373182 3 US-09-949-016-17371 Sequence 17371, A
c 38 15.8 75.2 373694 3 US-09-949-016-12062 Sequence 12062, A
c 39 15.8 75.2 678533 3 US-09-949-016-14577 Sequence 14577, A
c 40 15.8 75.2 678533 3 US-09-949-016-14578 Sequence 14578, A
c 41 15.8 75.2 784019 3 US-09-949-016-14033 Sequence 14033, A
c 42 15.8 75.2 828152 3 US-09-949-016-12777 Sequence 12777, A
c 43 15.4 73.3 1965 3 US-09-248-796A-5759 Sequence 5759, Ap
c 44 15.4 73.3 91772 3 US-09-949-016-15568 Sequence 15568, A
c 45 15.4 73.3 106199 3 US-09-949-016-12393 Sequence 12393, A
c 46 15.4 73.3 107980 3 US-09-949-016-14370 Sequence 14370, A
c 47 15.4 73.3 205183 3 US-09-949-016-17009 Sequence 17009, A
c 48 15.2 72.4 573 2 US-08-286-819A-20 Sequence 20, Appl
c 49 15.2 72.4 573 3 US-08-980-357-20 Sequence 20, Appl
c 50 15.2 72.4 885 3 US-09-357-375-20 Sequence 20, Appl
c 51 15.2 72.4 885 3 US-09-270-767-28376 Sequence 28376, A
c 52 15.2 72.4 1306 3 US-09-270-767-12594 Sequence 12594, A
c 53 15.2 72.4 7225 2 US-08-286-819A-15 Sequence 15, Appl
c 54 15.2 72.4 7225 3 US-08-980-357-15 Sequence 15, Appl
c 55 15.2 72.4 7225 3 US-09-357-375-15 Sequence 15, Appl
c 56 15.2 72.4 10851 2 US-08-286-819A-16 Sequence 16, Appl
c 57 15.2 72.4 10851 3 US-08-980-357-16 Sequence 16, Appl
c 58 15.2 72.4 10851 3 US-09-357-375-16 Sequence 16, Appl
c 59 15.2 72.4 29121 3 US-09-949-016-11748 Sequence 11748, A
c 60 15.2 72.4 29122 3 US-09-949-016-13591 Sequence 13591, A

```

#### ALIGNMENTS

RESULT 1  
US-10-081-923-2  
; Sequence 2, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-081-923-2

Query Match 100.0%; Score 21; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGCATGATGGGTATCTTCC 21  
|||||

Db 1 TGCATGATGGGTATCTTCC 21  
|||||

RESULT 2

US-10-081-923-12/c

```
; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 2
US-10-081-923-12

Query Match          100.0%; Score 21; DB 3; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 193 TGCATGTATGGGTATCTTCC 173

RESULT 3
US-10-081-923-14/c
; Sequence 14, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 10
US-10-081-923-14

Query Match          100.0%; Score 21; DB 3; Length 1378;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 4
US-10-081-923-7/c
; Sequence 7, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 1
US-10-081-923-7

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 210 TGCATGTATGGGTATCTTCC 190

; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 6
US-10-081-923-7

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 5
US-10-081-923-13/c
; Sequence 13, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 4
US-10-081-923-13

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 6
US-10-081-923-16/c
; Sequence 16, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 1
US-10-081-923-16

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   |||||
Db 209 TGCATGTATGGGTATCTTCC 189
```

```
RESULT 7
US-10-081-923-8/c
; Sequence 8, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 5
US-10-081-923-8
Query Match 100.0%; Score 21; DB 3; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 8
US-10-081-923-10/c
; Sequence 10, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 8
US-10-081-923-10
Query Match 100.0%; Score 21; DB 3; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 9
US-10-081-923-9/c
; Sequence 9, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 7
US-10-081-923-9
Query Match 100.0%; Score 21; DB 3; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 10
US-10-081-923-11/c
; Sequence 11, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 9
US-10-081-923-11
Query Match 100.0%; Score 21; DB 3; Length 1390;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 224 TGCATGTATGGTTATCTTCC 204

RESULT 11
US-10-081-923-15/c
; Sequence 15, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 3
US-10-081-923-15
Query Match 100.0%; Score 21; DB 3; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 210 TGCATGTATGGGTTATCTTCC 190

## RESULT 12

US-10-081-923-17/c  
; Sequence 17, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 1447  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: pIsI sequence from isolate no. 11  
US-10-081-923-17

Query Match 100.0%; Score 21; DB 3; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 210 TGCATGTATGGGTTATCTTCC 190

## RESULT 13

US-10-081-923-5/c  
; Sequence 5, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: pIsI sequence from Oklahoma University M1 strain  
; PUBLICATION INFORMATION:  
; AUTHORS: Ferretti et al.  
; JOURNAL: Proc. Natl. Acad. Sci. USA  
; VOLUME: 98  
; PAGES: 4658-4663  
; DATE: 2001-01-01  
US-10-081-923-5

Query Match 100.0%; Score 21; DB 3; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 377 TGCATGTATGGGTTATCTTCC 357

## RESULT 14

US-09-583-110-1191/c  
; Sequence 1191, Application US/09583110  
; Patent No. 669703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 1191  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1191

Query Match 92.4%; Score 19.4; DB 3; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
Db 377 TGCATGTATGGGTTATCTTCC 357

RESULT 15

US-09-107-433-1469/c  
; Sequence 1469, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1469:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1287 base pairs  
; TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1287  
SEQUENCE DESCRIPTION: SEQ ID NO: 1469;  
US-09-107-433-1469  
Query Match 92.4%; Score 19.4; DB 3; Length 1387;  
Best Local Similarity 95.2%; Pred. No. 3.4; Mismatches 1; Indels 0; Gaps 0;  
Matches 20; Conservative 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
431 TGCATGTATGGGTATCTTCC 411  
Db  
RESULT 16  
US-08-961-527-94/c  
Sequence 94, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8195 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-94  
Query Match 92.4%; Score 19.4; DB 3; Length 8195;  
Best Local Similarity 95.2%; Pred. No. 4.6; Mismatches 1; Indels 0; Gaps 0;  
Matches 20; Conservative 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
1384 TGCATGTATGGGTATCTTCC 1364  
Db

RESULT 17  
US-09-949-016-16988  
Sequence 16988, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16988  
LENGTH: 89210  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-16988  
Query Match 80.0%; Score 16.8; DB 3; Length 89210;  
Best Local Similarity 90.0%; Pred. No. 1.4e+02; Mismatches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 20  
70426 TGCATGTATGGGTATCTTCC 70445  
Db  
RESULT 18  
US-09-949-016-15890  
Sequence 15890, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15890  
LENGTH: 144922  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15890  
Query Match 80.0%; Score 16.8; DB 3; Length 144922;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02; Mismatches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 20  
130653 TGCATGTATGGGTATCTTCC 130672  
Db  
RESULT 19  
US-09-328-352-540  
Sequence 540, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:

```
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 540
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-540

Query Match          77.1%; Score 16.2; DB 3; Length 1059;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
    ||||| ||||| ||||| ||||| |||||
Db 561 TGCATTTATGGCTTATTTCC 581

RESULT 20
US-09-248-796A-2116/c
; Sequence 2116, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 2116
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-2116

Query Match          76.2%; Score 16; DB 3; Length 1356;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGTATGGGTTATCTTC 20
    ||||| ||||| ||||| |||||
Db 1258 TGTATGGGTTATCTTC 1243

RESULT 21
US-09-351-200-1/c
; Sequence 1, Application US/09351200
; Patent No. 6320033
; GENERAL INFORMATION:
; APPLICANT: BOURBONNAIS, Yves
; APPLICANT: LAMARE, Claude
; APPLICANT: DESLAURIERS, No. 6320033la
; TITLE OF INVENTION: CANDIDA ALBICANS GENE (CSA1) ENCODING A
; FILE REFERENCE: 6013-71"US" CC/
; CURRENT APPLICATION NUMBER: US/09/351,200
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: CA2,237,134
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4291
; TYPE: DNA

; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: gene
; LOCATION: (0)...(0)
US-09-351-200-1

Query Match          76.2%; Score 16; DB 3; Length 4291;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGTATGGGTTATCTTC 20
    ||||| ||||| ||||| |||||
Db 2186 TGTATGGGTTATCTTC 2171

RESULT 22
US-09-949-016-24376/c
; Sequence 24376, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24376
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24376

Query Match          75.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCATGTATGGGTTATCTT 19
    ||||| ||||| ||||| |||||
Db 28 TGCATATATAGCTTATCTT 10

RESULT 23
US-09-949-016-24696
; Sequence 24696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24696
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24696
```





```
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 784..7890
; US-08-198-446B-14

Query Match          75.2%; Score 15.8; DB 2; Length 8351;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTT 19
    ||||| ||| ||||| |||||
Db 7058 TGCATATATGGTATCTT 7040

RESULT 28
US-08-870-693-14/c
; Sequence 14, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Pilon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR110798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 784..7890
; US-08-870-693-14

Query Match          75.2%; Score 15.8; DB 2; Length 8351;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTT 19
    ||||| ||| ||||| |||||
Db 7058 TGCATATATGGTATCTT 7040

RESULT 29
US-09-949-016-14814
; Sequence 14814, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14814
; LENGTH: 9721
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14814

Query Match          75.2%; Score 15.8; DB 3; Length 9721;
Best Local Similarity 89.5%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTCC 21
    ||||| ||||| |||||
Db 2756 CATGTATGGGTAACTTTC 2774

RESULT 30
US-09-949-016-12076
; Sequence 12076, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12076
; LENGTH: 9722
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12076

Query Match          75.2%; Score 15.8; DB 3; Length 9722;
Best Local Similarity 89.5%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 CATGTATGGGTATCTTCC 21
      ||||| ||||| ||||| |||||
Db      2761 CATGTATGGGTAACTTC 2779

RESULT 31
US-09-949-016-16250/c
; Sequence 16250, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16250
; LENGTH: 26831
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16250

Query Match          75.2%; Score 15.8; DB 3; Length 26831;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTT 19
      ||||| ||||| ||||| |||||
Db      2875 TGCATGTATGGGTACCAT 2857

RESULT 32
US-09-949-016-16994/c
; Sequence 16994, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16994
; LENGTH: 36755
; TYPE: DNA

; ORGANISM: Human
US-09-949-016-16994

Query Match          75.2%; Score 15.8; DB 3; Length 36755;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTT 19
      ||||| ||||| ||||| |||||
Db      25570 TGCATGTATGAGTGATCTT 25552

RESULT 33
US-09-328-925-4/c
; Sequence 4, Application US/09328925
; Patent No. 6610906
; GENERAL INFORMATION:
; APPLICANT: Kurachi, Kotoku
; APPLICANT: Kurachi, Sumiko
; TITLE OF INVENTION: Nucleotide Sequences for Gene Regulation and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: UM-03603
; CURRENT APPLICATION NUMBER: US/09/328,925
; CURRENT FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 38059
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-925-4

Query Match          75.2%; Score 15.8; DB 3; Length 38059;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTT 19
      ||||| ||||| ||||| |||||
Db      26539 TGCATGTATGAGTGATCTT 26521

RESULT 34
US-09-949-016-15926
; Sequence 15926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15926
; LENGTH: 41435
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15926

Query Match          75.2%; Score 15.8; DB 3; Length 41435;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 20
      ||||| ||||| ||||| |||||
Db      40530 GCATATATGGGTATTTTC 40548
```

```
RESULT 35
US-09-949-016-13845
; Sequence 13845, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13845
; LENGTH: 152132
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13845

Query Match      75.2%; Score 15.8; DB 3; Length 152132;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      131531 TGAATGTATGGCTTATCTT 131549

RESULT 36
US-09-949-016-12371
; Sequence 12371, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12371
; LENGTH: 152145
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12371

Query Match      75.2%; Score 15.8; DB 3; Length 152145;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      131531 TGAATGTATGGCTTATCTT 131549

RESULT 37
US-09-949-016-17371/c
; Sequence 17371, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17371
; LENGTH: 373182
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(373182)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17371

Query Match      75.2%; Score 15.8; DB 3; Length 373182;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      66585 TGCATATATGAGTTATCTT 66567

RESULT 38
US-09-949-016-12062/c
; Sequence 12062, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12062
; LENGTH: 373694
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(373694)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12062

Query Match      75.2%; Score 15.8; DB 3; Length 373694;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      66640 TGCATATATGAGTTATCTT 66622
```

```
RESULT 39
US-09-949-016-14577
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577
Query Match 75.2%; Score 15.8; DB 3; Length 678533;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTCC 21
||| ||||| ||||| |||||
Db 590903 CATTATGGGTATCTTCC 590921

RESULT 40
US-09-949-016-14578
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14578
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578
Query Match 75.2%; Score 15.8; DB 3; Length 678533;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTCC 21
||| ||||| ||||| |||||
Db 590903 CATTATGGGTATCTTCC 590921

RESULT 41
US-09-949-016-14033/c
; Sequence 14033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14033
; LENGTH: 784019
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(784019)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14033
Query Match 75.2%; Score 15.8; DB 3; Length 784019;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTT 19
||||| ||||| ||||| |||||
Db 702418 TGCATGTATGTGTTCTT 702400

RESULT 42
US-09-949-016-12777/c
; Sequence 12777, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12777
; LENGTH: 828152
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(828152)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12777
Query Match 75.2%; Score 15.8; DB 3; Length 828152;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1  TGCATGATGGGTATCTT 19
Db      698551 TGCATGATGTTGTTCTT 698533

RESULT 43
US-09-248-796A-5759/c
; Sequence 5759, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5759
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5759

Query Match      73.3%; Score 15.4; DB 3; Length 1965;
Best Local Similarity 94.1%; Pred. No. 3.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  ATGTATGGGTATCTTC 20
Db      800 ATGTATGGTTATCTTC 784

RESULT 44
US-09-949-016-15568/c
; Sequence 15568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15568
; LENGTH: 91772
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(91772)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15568

Query Match      73.3%; Score 15.4; DB 3; Length 91772;
Best Local Similarity 94.1%; Pred. No. 6.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5  TGTATGGGTATCTTC 21
Db      64103 TGTATGGTTATCTTC 64087

RESULT 45
US-09-949-016-12393
; Sequence 12393, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12393
; LENGTH: 106199
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(106199)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12393

Query Match      73.3%; Score 15.4; DB 3; Length 106199;
Best Local Similarity 94.1%; Pred. No. 7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4  ATGTATGGGTATCTTC 20
Db      24733 AIGTTGGTTATCTTC 24749

RESULT 46
US-09-949-016-14370
; Sequence 14370, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14370
; LENGTH: 107980
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(107980)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14370

Query Match      73.3%; Score 15.4; DB 3; Length 107980;
Best Local Similarity 94.1%; Pred. No. 7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY      4 ATGATGGGTTATCTTC 20
      ||||| ||||| ||||| |||||
Db      74514 AAGTTTGGGTTATCTTC 74530

RESULT 47
US-09-949-016-17009/c
; Sequence 17009, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17009
; LENGTH: 205163
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17009

Query Match      73.3%; Score 15.4; DB 3; Length 205163;
Best Local Similarity 94.1%; Pred. No. 7.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATC 17
      ||||| ||||| ||||| |||||
Db      76139 TGCATGTATGGGTTAAC 76123

RESULT 48
US-08-286-819A-20/c
; Sequence 20, Application US/08286819A
; Patent No. 5871910
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN PARTICULAR
; IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,819A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5871910man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..573
US-08-286-819A-20

Query Match      72.4%; Score 15.2; DB 2; Length 573;
Best Local Similarity 85.0%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
      ||||| ||||| ||||| |||||
Db      317 TGGCTGTATGGATTATCTTC 298

RESULT 49
US-08-980-357-20/c
; Sequence 20, Application US/08980357
; Patent No. 6013508
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN PARTICULAR
; IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,357
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: US 08/286,819
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; APPLICATION DATA: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; APPLICATION NUMBER: FR 9013579
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 6013508man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..573
US-08-980-357-20

Query Match 72.4%; Score 15.2; DB 3; Length 573;
Best Local Similarity 85.0%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 317 TGGCTGTATGGATTATCTTC 298

RESULT 50
US-09-357-375-20/c
; Sequence 20, Application US/09357375
; Patent No. 6916906
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERIDES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,375
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,819
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; APPLICATION DATA: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; APPLICATION NUMBER: FR 9013579
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 6916906man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..573
US-09-357-375-20

Query Match 72.4%; Score 15.2; DB 3; Length 573;
Best Local Similarity 85.0%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 317 TGGCTGTATGGATTATCTTC 298

RESULT 51
US-09-270-767-28376
; Sequence 28376, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28376
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-28376

Query Match 72.4%; Score 15.2; DB 3; Length 885;
Best Local Similarity 85.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 698 TGGATATATCGGTTATCTTC 717

RESULT 52
US-09-270-767-12584
; Sequence 12584, Application US/09270767
```





/ TELEPHONE: (703) 413-3000  
/ TELEFAX: (703) 413-2220  
/ TELEX: 248855 OPAT UR  
/ INFORMATION FOR SEQ ID NO: 15:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 7225 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: unknown  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: DNA (genomic)  
US-08-980-357-15

Query Match 72.4%; Score 15.2; DB 3; Length 7225;  
Best Local Similarity 85.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
|| ||||| |||||  
Db 1048 TGGCTGTATGGATTATCTTC 1029

## RESULT 55

US-09-357-375-15/c  
/ Sequence 15, Application US/09357375  
/ Patent No. 6916906

/ GENERAL INFORMATION:  
/ APPLICANT: ARTHUR, MICHEL

/ APPLICANT: DUKTA-MALEN, SYLVIE  
/ APPLICANT: MOLINAS, CATHERINE

/ APPLICANT: COURVALIN, PATRICE  
/ TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

/ TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

/ TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

/ TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

/ NUMBER OF SEQUENCES: 54

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

/ ADDRESSEE: P.C.

/ STREET: 1755 S. Jefferson Davis Highway, Suite 400

/ CITY: Arlington

/ STATE: Virginia

/ COUNTRY: U.S.A.

/ ZIP: 22202

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patent In Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/09/357,375

/ FILING DATE:

/ CLASSIFICATION:

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/286,819

/ FILING DATE: 05-AUG-1994

/ APPLICATION NUMBER: US 08/174,682

/ FILING DATE: 28-DEC-1993

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 07/917,146

/ FILING DATE: 10-AUG-1992

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: PCT/FR/91/00855

/ FILING DATE: 29-OCT-1991

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: FR 9013579

/ FILING DATE: 31-OCT-1990

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Oblon, No. 6916906man F.

/ REGISTRATION NUMBER: 24,618

/ REFERENCE/DOCKET NUMBER: 660-060-0 PCT

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (703) 413-3000

/ TELEFAX: (703) 413-2220

/ TELEX: 248855 OPAT UR  
/ INFORMATION FOR SEQ ID NO: 15:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 7225 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: unknown  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: DNA (genomic)  
US-09-357-375-15

Query Match 72.4%; Score 15.2; DB 3; Length 7225;  
Best Local Similarity 85.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
|| ||||| |||||  
Db 1048 TGGCTGTATGGATTATCTTC 1029

## RESULT 56

US-08-286-819A-16/c

/ Sequence 16, Application US/08286819A

/ Patent No. 5871910

/ GENERAL INFORMATION:

/ APPLICANT: ARTHUR, MICHEL

/ APPLICANT: DUKTA-MALEN, SYLVIE

/ APPLICANT: MOLINAS, CATHERINE

/ APPLICANT: COURVALIN, PATRICE

/ TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

/ TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

/ TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

/ TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

/ NUMBER OF SEQUENCES: 54

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

/ ADDRESSEE: P.C.

/ STREET: 1755 S. Jefferson Davis Highway, Suite 400

/ CITY: Arlington

/ STATE: Virginia

/ COUNTRY: U.S.A.

/ ZIP: 22202

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patent In Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/286,819A

/ FILING DATE: 05-AUG-1994

/ CLASSIFICATION: 435

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/174,682

/ FILING DATE: 28-DEC-1993

/ CLASSIFICATION: 435

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 07/917,146

/ FILING DATE: 10-AUG-1992

/ CLASSIFICATION: 435

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: PCT/FR/91/00855

/ FILING DATE: 29-OCT-1991

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: FR 9013579

/ FILING DATE: 31-OCT-1990

/ CLASSIFICATION: 435

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Oblon, No. 5871910man F.

/ REGISTRATION NUMBER: 24,618

/ REFERENCE/DOCKET NUMBER: 660-060-0 PCT

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (703) 413-3000

/ TELEFAX: (703) 413-2220

/ TELEX: 248855 OPAT UR

;  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10851 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-286-819A-16

Query Match 72.4%; Score 15.2; DB 2; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
DB 3503 TGGCTGATGGATTATCTTC 3484

RESULT 57

US-08-980-357-16/c  
; Sequence 16, Application US/08980357  
; Patent No. 6013508

GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/980,357

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,819

FILING DATE: 05-AUG-1994

APPLICATION NUMBER: US 08/174,682

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,146

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6013508man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

;  
; LENGTH: 10851 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-980-357-16

Query Match 72.4%; Score 15.2; DB 3; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
DB 3503 TGGCTGATGGATTATCTTC 3484

RESULT 58

US-09-357-375-16/c  
; Sequence 16, Application US/09357375  
; Patent No. 6916906

GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/357,375

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,819

FILING DATE: 05-AUG-1994

APPLICATION NUMBER: US 08/174,682

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,146

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6916906man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 10851 base pairs

TYPE: nucleic acid

;  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-09-357-375-16

Query Match 72.4%; Score 15.2; DB 3; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
||| ||||| ||||| |||||  
Db 3503 TGGCTGTATGGATTATCTTC 3484

RESULT 59  
US-09-949-016-11748  
; Sequence 11748, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11748  
; LENGTH: 29121  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11748

Query Match 72.4%; Score 15.2; DB 3; Length 29121;  
Best Local Similarity 85.0%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
||| ||||| ||||| |||||  
Db 13951 GCATGTATTGTTATGTTCC 13970

RESULT 60  
US-09-949-016-13591  
; Sequence 13591, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13591  
; LENGTH: 29122  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13591

Query Match 72.4%; Score 15.2; DB 3; Length 29122;  
Best Local Similarity 85.0%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
||| ||||| ||||| |||||  
Db 13951 GCATGTATTGTTATGTTCC 13970

Search completed: January 28, 2006, 01:18:37  
Job time : 66.9888 secs

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 231.809 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23  
Sequence: 1 tgagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 41533918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listings first 60 summaries

Database : Published Applications NA New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US12\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US13\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US14\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US15\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US16\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	81.7	201	7	US-10-995-561-48428
2	18.8	81.7	81210	7	US-10-995-561-13295
3	18.8	81.7	171732	8	US-11-121-086-98
4	17.2	74.8	24	7	US-10-310-914A-636042
5	17.2	74.8	886	5	US-09-978-360A-196
6	17.2	74.8	17517	8	US-11-136-527-3650
7	16.8	73.0	769	7	US-10-750-185-43151
8	16.8	73.0	769	7	US-10-750-623-43151
9	16.8	73.0	150450	8	US-11-112-908-54
10	16.8	73.0	191343	8	US-11-112-908-53
11	16.6	72.2	201	8	US-11-124-368A-20510
12	16.6	72.2	201	8	US-11-124-368A-20540
13	16.6	72.2	600	7	US-10-750-185-2581
14	16.6	72.2	600	7	US-10-750-185-2668
15	16.6	72.2	600	7	US-10-750-623-2581
16	16.6	72.2	600	7	US-10-750-623-2668
17	16.6	72.2	600	8	US-11-136-527-7331
18	16.6	72.2	933	7	US-10-750-185-40685
19	16.6	72.2	933	7	US-10-750-623-40685
20	16.6	72.2	1962	8	US-11-136-527-3235
21	16.6	72.2	2200	7	US-10-750-185-59759
22	16.6	72.2	2200	7	US-10-750-623-59759

23	16.6	72.2	3257	8	US-11-128-061-3404	Sequence 3404, Ap
24	16.6	72.2	3257	8	US-11-128-049-3404	Sequence 3404, Ap
25	16.6	72.2	41081	8	US-11-124-368A-2931	Sequence 2931, Ap
26	16.6	72.2	165627	8	US-11-121-086-89	Sequence 89, Appl
27	16.4	71.3	201	7	US-10-995-561-59465	Sequence 59465, A
28	16.4	71.3	600	8	US-11-136-527-6022	Sequence 6022, Ap
29	16.4	71.3	2391	8	US-11-136-527-1926	Sequence 1926, Ap
30	16.4	71.3	83528	7	US-10-995-561-13343	Sequence 13343, A
31	16.4	71.3	162085	8	US-11-121-086-7	Sequence 7, Appl1
32	16.2	70.4	21	7	US-10-310-914A-635991	Sequence 635991, A
33	16.2	70.4	25	8	US-11-121-849-93224	Sequence 93224, A
34	16.2	70.4	25	8	US-11-121-849-94187	Sequence 94187, A
35	16.2	70.4	201	7	US-10-995-561-27893	Sequence 27893, A
36	16.2	70.4	201	7	US-10-995-561-72197	Sequence 72197, A
37	16.2	70.4	482	8	US-11-128-061-2983	Sequence 2983, Ap
38	16.2	70.4	482	8	US-11-128-061-6625	Sequence 6625, Ap
39	16.2	70.4	482	8	US-11-128-049-2983	Sequence 2983, Ap
40	16.2	70.4	482	8	US-11-128-049-6625	Sequence 6625, Ap
41	16.2	70.4	984	7	US-10-750-185-29356	Sequence 29356, A
42	16.2	70.4	984	7	US-10-750-623-29356	Sequence 29356, A
43	16.2	70.4	7320	7	US-10-486-968-1	Sequence 1, Appl1
44	16.2	70.4	47622	7	US-10-995-561-13431	Sequence 13431, A
45	16.2	70.4	53328	7	US-10-995-561-13431	Sequence 13431, A
46	16	69.6	22	7	US-10-310-914A-1196435	Sequence 1196435, A
47	16	69.6	24	7	US-10-310-914A-1196435	Sequence 1196435, A
48	15.8	68.7	28	7	US-10-310-914A-406624	Sequence 406624, A
49	15.8	68.7	1041	7	US-10-750-185-63692	Sequence 63692, A
50	15.8	68.7	1041	7	US-10-750-623-63692	Sequence 63692, A
51	15.8	68.7	1066	7	US-10-750-185-26352	Sequence 26352, A
52	15.8	68.7	1066	7	US-10-750-623-26352	Sequence 26352, A
53	15.8	68.7	73404	8	US-11-124-368A-2914	Sequence 2914, Ap
54	15.8	68.7	190376	7	US-10-661-966-1	Sequence 1, Appl1
55	15.8	68.7	191350	7	US-10-857-780-4	Sequence 4, Appl1
56	15.8	68.7	1691140	8	US-11-091-018-1	Sequence 1, Appl1
57	15.6	67.8	22	7	US-10-310-914A-156548	Sequence 156548, A
58	15.6	67.8	25	8	US-11-136-527-79009	Sequence 79009, A
59	15.6	67.8	25	8	US-11-136-527-79014	Sequence 79014, A
60	15.6	67.8	25	8	US-11-136-527-278745	Sequence 278745, A

ALIGNMENTS

RESULT 1

US-10-995-561-48428/c

; Sequence 48428, Application US/109955561

; Publication No. US20050272054A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

; TITLE OF INVENTION: DETECTION AND USES THEREOF

; FILE REFERENCE: CLO01559

; CURRENT APPLICATION NUMBER: US/10/995,561

; CURRENT FILING DATE: 2004-11-24

; NUMBER OF SEQ ID NOS: 85702

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 48428

; LENGTH: 201

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-995-561-48428

Query Match 81.7%; Score 18.8; DB 7; Length 201;

Best Local Similarity 90.9%; Pred. No. 10;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGCTTAG 23

|||||

Db 25 GAGAGGCAGTAGAAGCTTAG 4

RESULT 2

```
US-10-995-561-13295/c
; Sequence 13295, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13295
; LENGTH: 81210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13295
Query Match      81.7%; Score 18.8; DB 7; Length 81210;
Best Local Similarity 90.9%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GAGAGGCAGTAGAAGCTTAG 23
      ||||| ||||| ||||| |||||
Db      525 GAGAGGAAGTAGAAGCCTAG 504

RESULT 3
US-11-121-086-98/c
; Sequence 98, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 98
; LENGTH: 171732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-98
Query Match      81.7%; Score 18.8; DB 8; Length 171732;
Best Local Similarity 90.9%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GAGAGGCAGTAGAAGCTTAG 23
      ||||| ||||| ||||| |||||
Db      131125 GAGAGGAAGTAGAAGCCTAG 131104

RESULT 4
US-10-310-914A-636042
; Sequence 636042, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 636042
```

```
; LENGTH: 24
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-636042
Query Match      74.8%; Score 17.2; DB 7; Length 24;
Best Local Similarity 81.8%; Pred. No. 40;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAAGGCAGTAGAAAAGCTTAG 23
      ||||| ||||| ||||| |||||
Db      2 GAGAAGGCAGCAGAAAAGAUGAG 23

RESULT 5
US-09-978-360A-196/c
; Sequence 196, Application US/09978360A
; Publication No. US20060009633A9
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclet, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; APPLICANT: Clusel, Catherine
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
; FILE REFERENCE: 56.US4.CIP
; CURRENT APPLICATION NUMBER: US/09/978,360A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: US 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: US 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: US 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: US 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: US 60/099,273
; PRIOR FILING DATE: -09-04
; PRIOR APPLICATION NUMBER: US 09/191,997
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: US 09/215,435
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: PCT/IB98/02122
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: US 09/247,155
; PRIOR FILING DATE: 1999-02-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 810
; SOFTWARE: Patent.pm
; SEQ ID NO 196
; LENGTH: 886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..381
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 31..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.4
; OTHER INFORMATION: seq AFVIACVLSLIST/IY
; FEATURE:
; NAME/KEY: polyA site
; LOCATION: 875..886
US-09-978-360A-196
Query Match      74.8%; Score 17.2; DB 5; Length 886;
Best Local Similarity 86.4%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 2 GAGAGGCGAGTAAAGCTTAG 23  
Db 473 GAGAGGCGAGTAAAGAGAG 452

RESULT 6  
US-11-136-527-3650/c  
; Sequence 3650, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3650  
; LENGTH: 17517  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-3650

Query Match 74.8%; Score 17.2; DB 8; Length 17517;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAAAGCTTAG 23  
Db 13480 GAGAGGCGAGGCAAGCTGAG 13459

RESULT 7  
US-10-750-185-43151  
; Sequence 43151, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 43151  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Bovine 19866881328491  
US-10-750-185-43151

Query Match 73.0%; Score 16.8; DB 7; Length 769;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAAAGCTT 21  
Db 660 GAGATGCGAGCAAGCTT 679

RESULT 8  
US-10-750-623-43151  
; Sequence 43151, Application US/10750623

; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 43151  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Bovine 19866881328491  
US-10-750-623-43151

Query Match 73.0%; Score 16.8; DB 7; Length 769;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAAAGCTT 21  
Db 660 GAGATGCGAGCAAGCTT 679

RESULT 9  
US-11-112-908-54  
; Sequence 54, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 54  
; LENGTH: 150450  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-54

Query Match 73.0%; Score 16.8; DB 8; Length 150450;  
Best Local Similarity 90.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGGCGAGTAAAGCTTAG 23  
Db 145183 GAGACAGGAGAAAGCTTAG 145202

RESULT 10  
US-11-112-908-53  
; Sequence 53, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:

```
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 191343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-53

Query Match      73.0%; Score 16.8; DB 8; Length 191343;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAAGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 64425 GAAGACAGGAGAAAGCTTAG 64444

RESULT 11
US-11-124-368A-20510
; Sequence 20510, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

RESULT 12
US-11-124-368A-20540
; Sequence 20540, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 191343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-53

Query Match      73.0%; Score 16.8; DB 8; Length 191343;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAAGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 64425 GAAGACAGGAGAAAGCTTAG 64444

RESULT 11
US-11-124-368A-20510
; Sequence 20510, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

RESULT 12
US-11-124-368A-20540
; Sequence 20540, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20540
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20540

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 105 TGGGAAGGAAGGAGAAAGCTTAG 127

RESULT 13
US-10-750-185-2581
; Sequence 2581, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2581
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine
MMBT13689
US-10-750-185-2581

Query Match      72.2%; Score 16.6; DB 7; Length 600;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 425 TGTGAAGGCAGGAGAAAAGCTGAG 447

RESULT 14
US-10-750-185-2668
; Sequence 2668, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```



;; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

;; FILE REFERENCE: MM1100-2

;; CURRENT APPLICATION NUMBER: US/10/750,185

;; CURRENT FILING DATE: 2003-12-31

;; PRIOR APPLICATION NUMBER: US 60/437,482

;; PRIOR FILING DATE: 2002-12-31

;; NUMBER OF SEQ ID NOS: 64922

;; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 2668

;; LENGTH: 600

;; TYPE: DNA

;; ORGANISM: Bovine MMBT08581

US-10-750-185-2668

Query Match 72.2%; Score 16.6; DB 7; Length 600;

Best Local Similarity 82.6%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23

Db 443 TGTGAAGGCAGGAAAAGCTGAG 465

RESULT 15

US-10-750-623-2581

;; Sequence 2581, Application US/10750623

;; Publication No. US20050287531A1

;; GENERAL INFORMATION:

;; APPLICANT: MMI GENOMICS, INC.

;; APPLICANT: DENISE, Sue K.

;; APPLICANT: KERR, Richard

;; APPLICANT: ROSENFELD, David

;; APPLICANT: HOLM, Tom

;; APPLICANT: BATES, Stephen

;; APPLICANT: FANTIN, Dennis

;; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

;; FILE REFERENCE: MM1100-1

;; CURRENT APPLICATION NUMBER: US/10/750,623

;; CURRENT FILING DATE: 2003-12-31

;; PRIOR APPLICATION NUMBER: US 60/437,482

;; PRIOR FILING DATE: 2002-12-31

;; NUMBER OF SEQ ID NOS: 64922

;; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 2581

;; LENGTH: 600

;; TYPE: DNA

;; ORGANISM: Bovine MMBT13689

US-10-750-623-2581

Query Match 72.2%; Score 16.6; DB 7; Length 600;

Best Local Similarity 82.6%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23

Db 425 TGTGAAGGCAGGAAAAGCTGAG 447

RESULT 16

US-10-750-623-2668

;; Sequence 2668, Application US/10750623

;; Publication No. US20050287531A1

;; GENERAL INFORMATION:

;; APPLICANT: MMI GENOMICS, INC.

;; APPLICANT: DENISE, Sue K.

;; APPLICANT: KERR, Richard

;; APPLICANT: ROSENFELD, David

;; APPLICANT: HOLM, Tom

;; APPLICANT: BATES, Stephen

;; APPLICANT: FANTIN, Dennis

;; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

;; FILE REFERENCE: MM1100-1

;; CURRENT APPLICATION NUMBER: US/10/750,623

;; CURRENT FILING DATE: 2003-12-31

;; PRIOR APPLICATION NUMBER: US 60/437,482

;; PRIOR FILING DATE: 2002-12-31

;; NUMBER OF SEQ ID NOS: 64922

;; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 2668

;; LENGTH: 600

;; TYPE: DNA

;; ORGANISM: Bovine MMBT08581

US-10-750-623-2668

Query Match 72.2%; Score 16.6; DB 7; Length 600;

Best Local Similarity 82.6%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23

Db 443 TGTGAAGGCAGGAAAAGCTGAG 465

RESULT 17

US-11-136-527-7331

;; Sequence 7331, Application US/111136527

;; Publication No. US20050287570A1

;; GENERAL INFORMATION:

;; APPLICANT: Wyeth

;; APPLICANT: Mounts, William M

;; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes

;; FILE REFERENCE: 031896-041000 (AM101086)

;; CURRENT APPLICATION NUMBER: US/11/136,527

;; CURRENT FILING DATE: 2005-05-25

;; PRIOR APPLICATION NUMBER: US 60/574,294

;; PRIOR FILING DATE: 2005-05-26

;; NUMBER OF SEQ ID NOS: 362830

;; SOFTWARE: PatentIn version 3.2

;; SEQ ID NO 7331

;; LENGTH: 600

;; TYPE: DNA

;; ORGANISM: Rattus norvegicus

US-11-136-527-7331

Query Match 72.2%; Score 16.6; DB 8; Length 600;

Best Local Similarity 82.6%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23

Db 484 TGAGAGGTAGCAGCAAGCTGAG 506

RESULT 18

US-10-750-185-40685/c

;; Sequence 40685, Application US/10750185

;; Publication No. US20050260603A1

;; GENERAL INFORMATION:

;; APPLICANT: MMI GENOMICS, INC.

;; APPLICANT: DENISE, Sue K.

;; APPLICANT: KERR, Richard

;; APPLICANT: ROSENFELD, David

;; APPLICANT: HOLM, Tom

;; APPLICANT: BATES, Stephen

;; APPLICANT: FANTIN, Dennis

;; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

;; FILE REFERENCE: MM1100-2

;; CURRENT APPLICATION NUMBER: US/10/750,185

;; CURRENT FILING DATE: 2003-12-31

;; PRIOR APPLICATION NUMBER: US 60/437,482

;; PRIOR FILING DATE: 2002-12-31

;; NUMBER OF SEQ ID NOS: 64922

;; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 40685

;; LENGTH: 933

;; TYPE: DNA

```

; ORGANISM: Bovine 19866880900388
US-10-750-185-40685

Query Match      72.2%; Score 16.6; DB 7; Length 933;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 867 TGACACGGAAGAAGAAGCTTAG 845

RESULT 19
US-10-750-623-40685/c
; Sequence 40685, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40685
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Bovine 19866880900388
US-10-750-623-40685

Query Match      72.2%; Score 16.6; DB 7; Length 933;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 867 TGACACGGAAGAAGAAGCTTAG 845

RESULT 20
US-11-136-527-3235
; Sequence 3235, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3235
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-3235

Query Match      72.2%; Score 16.6; DB 8; Length 1962;
Best Local Similarity 82.6%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23

```

```

Db 1846 TGAGAAGGTAGCAGCAAGCTGAG 1868
    ||||| ||||| ||||| ||||| |||||

RESULT 21
US-10-750-185-59759
; Sequence 59759, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59759
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Bovine 19866880587237
US-10-750-185-59759

Query Match      72.2%; Score 16.6; DB 7; Length 2200;
Best Local Similarity 82.6%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 1837 TGATGAGGCAGTAGAAGGCTAG 1859

RESULT 22
US-10-750-623-59759
; Sequence 59759, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59759
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Bovine 19866880587237
US-10-750-623-59759

Query Match      72.2%; Score 16.6; DB 7; Length 2200;
Best Local Similarity 82.6%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 1837 TGATGAGGCAGTAGAAGGCTAG 1859

```

```
RESULT 23
US-11-128-061-3404
; Sequence 3404, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; PRIOR FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3404
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1066)..(1086)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2121)..(2140)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-3404

Query Match 72.2%; Score 16.6; DB 8; Length 3257;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCAGAAGCGCAGTAGAAGCTTAG 23
Db 1585 TCAGAAGCCAGCAGAAAGCTCAG 1607

RESULT 24
US-11-128-049-3404
; Sequence 3404, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3404
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1066)..(1086)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2121)..(2140)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-3404

Query Match 72.2%; Score 16.6; DB 8; Length 3257;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCAGAAGCGCAGTAGAAGCTTAG 23
Db 1585 TCAGAAGCCAGCAGAAAGCTCAG 1607

RESULT 25
US-11-124-368A-2931
; Sequence 2931, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2931
; LENGTH: 41081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2931

Query Match 72.2%; Score 16.6; DB 8; Length 41081;
Best Local Similarity 82.6%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCAGAAGCGCAGTAGAAGCTTAG 23
Db 27929 TCGGAAGGAAGGAGAAAGCTTAG 27951

RESULT 26
US-11-121-086-89
; Sequence 89, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 89
; LENGTH: 165627
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-89
```

```
Query Match          72.2%; Score 16.6; DB 8; Length 165627;
Best Local Similarity 82.6%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 101723 TAAGAAGGCAGCAGACATCTTAG 101745

RESULT 27
US-10-995-561-59465/c
; Sequence 59465, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59465
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-59465

Query Match          71.3%; Score 16.4; DB 7; Length 201;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTA 22
   ||||| ||||| ||||| |||||
Db 119 AGCAGGCAGTAGAAGATYA 100

RESULT 28
US-11-136-527-6022/c
; Sequence 6022, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6022
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6022

Query Match          71.3%; Score 16.4; DB 8; Length 600;
Best Local Similarity 77.3%; Pred. No. 1.6e+02;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 145 GAGAAGGCCCAASAARGCTTAG 124

RESULT 29
US-11-136-527-1926/c
; Sequence 1926, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1926
; LENGTH: 2391
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-1926

Query Match          71.3%; Score 16.4; DB 8; Length 2391;
Best Local Similarity 77.3%; Pred. No. 2.1e+02;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 1936 GAGAAGGCCCAASAARGCTTAG 1915

RESULT 30
US-10-995-561-13343/c
; Sequence 13343, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13343
; LENGTH: 83528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13343

Query Match          71.3%; Score 16.4; DB 7; Length 83528;
Best Local Similarity 85.0%; Pred. No. 3.8e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTA 22
   ||||| ||||| ||||| |||||
Db 16778 AGCAGGCAGTAGAAGATYA 16759

RESULT 31
US-11-121-086-7/c
; Sequence 7, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 162085
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

US-11-121-086-7
Query Match          71.3%; Score 16.4; DB 8; Length 162085;
Best Local Similarity 94.4%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTT 21
      |||||
Db      3544 GAAGGCAGTAGAAGCTT 3527

RESULT 32
US-10-310-914A-635991
; Sequence 635991, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; FILE REFERENCE: 6087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 635991
; LENGTH: 21
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-635991

Query Match          70.4%; Score 16.2; DB 7; Length 21;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAGCTTAG 23
      |||||
Db      1 AGAAGGCAGCAGAAAGUAG 21

RESULT 33
US-11-121-849-93224
; Sequence 93224, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 93224
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-93224

Query Match          70.4%; Score 16.2; DB 8; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
      |||||
Db      3 TGAGAAGGCAGTAGAAGCTT 23

RESULT 34
US-11-121-849-94187
; Sequence 94187, Application US/11121849

```

```

; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 94187
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-94187

Query Match          70.4%; Score 16.2; DB 8; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
      |||||
Db      3 TGAGAAGGCAGTAGAAGCTT 23

RESULT 35
US-10-995-561-27893
; Sequence 27893, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27893
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-27893

Query Match          70.4%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAGCTTAG 23
      |||||
Db      105 AGCAGGCTGTAGAGAGCTTAG 125

RESULT 36
US-10-995-561-72197
; Sequence 72197, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72197
; LENGTH: 201
; TYPE: DNA

```

```
; ORGANISM: Homo sapiens
US-10-995-561-72197

Query Match      70.4%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAGCTTAG 23
Db      105 AGCAGGCTGTAGAGAGCTTAG 125

RESULT 37
US-11-128-061-2983
; Sequence 2983, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2983
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-2983

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
Db      5 TGAGAAGGCAGTGCAAAAGCCT 25

RESULT 38
US-11-128-061-6625
; Sequence 6625, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
```

```
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6625
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-6625

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
Db      5 TGAGAAGGCAGTGCAAAAGCCT 25

RESULT 39
US-11-128-049-2983
; Sequence 2983, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2983
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-2983

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
Db      5 TGAGAAGGCAGTGCAAAAGCCT 25

RESULT 40
US-11-128-049-6625
; Sequence 6625, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
```

```

; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6625
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-6625

Query Match 70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21
   ||||| ||||| |||||
Db 5 TGAGAGGCGAGTAGAAGCCT 25

RESULT 41
US-10-750-185-29356/c
; Sequence 29356, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29356
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Bovine 19866881161138
US-10-750-185-29356

Query Match 70.4%; Score 16.2; DB 7; Length 984;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| |||||
Db 918 AGAAGGCGAGTAGAAGCTTAG 898

RESULT 42
US-10-750-623-29356/c
; Sequence 29356, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29356
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Bovine 19866881161138
US-10-750-623-29356

Query Match 70.4%; Score 16.2; DB 7; Length 984;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| |||||
Db 918 AGAAGGCGAGTAGAAGCTTAG 898

RESULT 43
US-10-486-968-1
; Sequence 1, Application US/10486968
; Publication No. US20050250205A1
; GENERAL INFORMATION:
; APPLICANT: Genoplate-Valor S.A.S
; TITLE OF INVENTION: Use of associations between at least one nucleic sequence
; TITLE OF INVENTION: Polymorphism of the SH2 gene and at least one seed
; FILE REFERENCE: SH2
; CURRENT APPLICATION NUMBER: US/10/486,968
; CURRENT FILING DATE: 2004-02-17
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7320
; TYPE: DNA
; ORGANISM: Zea mays
US-10-486-968-1

Query Match 70.4%; Score 16.2; DB 7; Length 7320;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| |||||
Db 1647 AGAAGGCGAGTAGAAGCTTTG 1667

RESULT 44
US-10-995-561-13431
; Sequence 13431, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 47622
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-995-561-13431

Query Match      70.4%; Score 16.2; DB 7; Length 47622;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
Db      38922 AGCAGGCTGTAGAGAGCTTAG 38942

RESULT 45
US-10-995-561-13243
; Sequence 13243, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13243
; LENGTH: 53328
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13243

Query Match      70.4%; Score 16.2; DB 7; Length 53328;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
Db      614 AGCAGGCTGTAGAGAGCTTAG 634

RESULT 46
US-10-914A-1196483/c
; Sequence 1196483, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1196483
; LENGTH: 22
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-1196483

Query Match      69.6%; Score 16; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAG 18
Db      22 AGAAGGCAGTAGAAAG 7

RESULT 47
US-10-310-914A-1196435/c
; Sequence 1196435, Application US/10310914A
```

```
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1196435
; LENGTH: 24
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-1196435

Query Match      69.6%; Score 16; DB 7; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAG 18
Db      18 AGAAGGCAGTAGAAAG 3

RESULT 48
US-10-310-914A-406624
; Sequence 406624, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 406624
; LENGTH: 28
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-406624

Query Match      68.7%; Score 15.8; DB 7; Length 28;
Best Local Similarity 73.7%; Pred. No. 1.8e+02;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTT 21
Db      6 AGAAGGCAGUAUAAACCUU 24

RESULT 49
US-10-750-185-63692
; Sequence 63692, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```



```

; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 26352

; GENERAL INFORMATION:
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 2112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2914
; LENGTH: 73404
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-124-368A-2914

```

```
Query Match      68.7%; Score 15.8; DB 8; Length 73404;
Best Local Similarity 89.5%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGCAGTAGAAGC 19
    ||||| ||||| ||||| |||||
Db 9784 TGAGAGGCGCAGTAGAAGC 9802

RESULT 54
US-10-661-966-1
; Sequence 1, Application US/10661966
; Publication No. US2005027118A1
; GENERAL INFORMATION:
; APPLICANT: Roth, Richard B.
; APPLICANT: Nelson, Matthew Roberts
; APPLICANT: Braun, Andreas
; TITLE OF INVENTION: METHODS FOR IDENTIFYING SUBJECTS AT RISK
; TITLE OF INVENTION: OF MELANOMA AND TREATMENTS THEREOF
; FILE REFERENCE: 524592003800
; CURRENT APPLICATION NUMBER: US/10/661,966
; CURRENT FILING DATE: 2003-09-11
; PRIOR APPLICATION NUMBER: 60/410,595
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 60/422,344
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 190276
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-661-966-1

Query Match      68.7%; Score 15.8; DB 7; Length 190276;
Best Local Similarity 89.5%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCGCAGTAGAAGCTT 21
    ||||| ||||| ||||| |||||
Db 113905 AGAAGGCGCAGTAGAAGCTT 113923

RESULT 55
US-10-857-780-4/c
; Sequence 4, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 191350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (16914)..(16914)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19266)..(19266)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (26334)..(26334)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (34472)..(34472)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (38628)..(38628)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40555)..(40555)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (57355)..(57355)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (61207)..(61207)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (64980)..(64980)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (64987)..(64987)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81452)..(81452)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (95893)..(95893)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103359)..(103359)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117565)..(117565)
; OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-4

Query Match      68.7%; Score 15.8; DB 7; Length 191350;
Best Local Similarity 81.0%; Pred. No. 7.7e+02;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGCAGTAGAAGCTT 21
    ||||| ||||| ||||| |||||
Db 17982 TGAGAGGCGCAGTAGAAGCTT 17962

RESULT 56
US-11-091-018-1
; Sequence 1, Application US/11091018
; Publication No. US20050287551A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Thorleifsson, Gudmar
```

```

; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; FILE REFERENCE: 2345.2010-016
; CURRENT APPLICATION NUMBER: US/11/091,018
; CURRENT FILING DATE: 2005-03-25
; PRIOR APPLICATION NUMBER: PCT/US03/29906
; PRIOR FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/419,723
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/650,120
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691140
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1691140)
; OTHER INFORMATION: n=A,T,C or G
US-11-091-018-1

Query Match      68.7%; Score 15.8; DB 8; Length 1691140;
Best Local Similarity 89.5%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5  AAGCAGTACGAGAAAGCTTAG 23
Db      798654  AAGCAGTACGAGTACTCAG 798672

RESULT 57
US-10-310-914A-156548
; Sequence 156548, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CFUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 156548
; LENGTH: 22
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-156548

Query Match      67.8%; Score 15.6; DB 7; Length 22;
Best Local Similarity 68.2%; Pred. No. 2.2e+02;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTACGAGAAAGCTTA 22
Db      1  UGAGAAGGCAGTACGAGAAAGCTTA 22

RESULT 58
US-11-136-527-79009
; Sequence 79009, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79009
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-79009

Query Match      67.8%; Score 15.6; DB 8; Length 25;
Best Local Similarity 81.8%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2  GAGAAGGCAGTACGAGAAAGCTTAG 23
Db      1  GATCAGGCAGAGAGAGGCTTAG 22

RESULT 59
US-11-136-527-79014
; Sequence 79014, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79014
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-79014

Query Match      67.8%; Score 15.6; DB 8; Length 25;
Best Local Similarity 81.8%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2  GAGAAGGCAGTACGAGAAAGCTTAG 23
Db      2  GATCAGGCAGAGAGAGGCTTAG 23

RESULT 60
US-11-136-527-278745
; Sequence 278745, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830

```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 278745
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-278745

Query Match      67.8%; Score 15.6; DB 8; Length 25;
Best Local Similarity 81.8%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 GAGAGGCAGTAGAAGCTTAG 23
      || ||||| ||||| |||||
Db       4 GATCAGGCAGAGAGGCTTAG 25

Search completed: January 28, 2006, 02:00:36
Job time : 240.059 secs
```

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 399.011 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23

Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	9	US-10-716-005-1
2	19.8	86.1	1378	7	US-10-465-205-14
3	19.8	86.1	1379	7	US-10-465-205-7
4	19.8	86.1	1379	7	US-10-465-205-13
5	19.8	86.1	1379	7	US-10-465-205-16
6	19.8	86.1	1384	7	US-10-465-205-8
7	19.8	86.1	1384	7	US-10-465-205-10
8	19.8	86.1	1385	7	US-10-465-205-9
9	19.8	86.1	1390	7	US-10-465-205-11
10	19.8	86.1	1393	7	US-10-465-205-15
11	19.8	86.1	1447	7	US-10-465-205-17
12	19.8	86.1	1734	7	US-10-282-122A-38442
13	19.8	86.1	1803	7	US-10-465-205-5
14	18.8	81.7	81210	7	US-10-741-601-5664
15	18.8	81.7	81210	8	US-10-741-600-17691
16	18.2	79.1	720	9	US-10-504-582-15
17	18.2	79.1	160820	8	US-10-473-939-5
18	17.8	77.4	522	4	US-09-525-065A-739139
19	17.8	77.4	525	4	US-09-525-065A-739199
20	17.8	77.4	545	4	US-09-525-065A-728631
21	17.8	77.4	624	4	US-09-525-065A-69405
22	17.8	77.4	863	4	US-09-525-065A-8982
23	17.8	77.4	1206	4	US-09-525-065A-709349

c	24	17.8	77.4	1206	4	US-09-925-065A-709350	Sequence 709350,
	25	17.8	77.4	1253	4	US-09-925-065A-62246	Sequence 62246, A
	26	17.8	77.4	1714	7	US-10-425-114-24004	Sequence 24004, A
c	27	17.8	77.4	1940	5	US-10-027-632-97987	Sequence 97987, A
c	28	17.8	77.4	1940	6	US-10-027-632-97987	Sequence 97987, A
	29	17.8	77.4	1991	8	US-10-425-115-112272	Sequence 112272, A
	30	17.8	77.4	2478	7	US-10-437-963-95439	Sequence 95439, A
c	31	17.8	77.4	2533	6	US-10-108-260A-1480	Sequence 1480, Ap
c	32	17.8	77.4	3279	5	US-10-027-632-114128	Sequence 114128,
c	33	17.8	77.4	3279	5	US-10-027-632-114129	Sequence 114129,
c	34	17.8	77.4	3279	6	US-10-027-632-114128	Sequence 114128,
c	35	17.8	77.4	3279	6	US-10-027-632-114129	Sequence 114129,
c	36	17.8	77.4	6805	10	US-11-097-143-16906	Sequence 16906, A
c	37	17.8	77.4	17391	7	US-10-741-601-5677	Sequence 5677, Ap
c	38	17.8	77.4	48265	7	US-10-741-601-5677	Sequence 5677, Ap
c	39	17.8	77.4	104399	7	US-10-322-281-123	Sequence 123, App
	40	17.4	75.7	21	7	US-10-465-205-1	Sequence 1, Appl
c	41	17.4	75.7	309	7	US-10-469-285-225	Sequence 225, App
c	42	17.4	75.7	322	3	US-09-917-800A-1065	Sequence 1065, Ap
c	43	17.4	75.7	569	4	US-09-925-065A-560212	Sequence 560212,
	44	17.4	75.7	2151	9	US-10-764-420-2470	Sequence 2470, Ap
	45	17.4	75.7	24789	10	US-11-097-143-40201	Sequence 40201, A
	46	17.2	74.8	300	7	US-10-242-535A-45247	Sequence 45247, A
c	47	17.2	74.8	300	7	US-10-085-783A-45247	Sequence 45247, A
	48	17.2	74.8	494	5	US-10-063-254-187	Sequence 187, App
	49	17.2	74.8	519	3	US-09-917-800A-436	Sequence 436, App
	50	17.2	74.8	546	5	US-10-027-632-237122	Sequence 237122,
	51	17.2	74.8	546	6	US-10-027-632-237122	Sequence 237122,
	52	17.2	74.8	568	5	US-10-027-632-49562	Sequence 49562, A
	53	17.2	74.8	568	5	US-10-027-632-49563	Sequence 49563, A
	54	17.2	74.8	568	5	US-10-027-632-49564	Sequence 49564, A
	55	17.2	74.8	568	6	US-10-027-632-49562	Sequence 49562, A
	56	17.2	74.8	568	6	US-10-027-632-49563	Sequence 49563, A
	57	17.2	74.8	568	6	US-10-027-632-49564	Sequence 49564, A
	58	17.2	74.8	583	4	US-09-925-065A-34	Sequence 34, Appl
	59	17.2	74.8	600	9	US-10-972-079-87189	Sequence 87189, A
	60	17.2	74.8	600	9	US-10-972-079-87190	Sequence 87190, A

ALIGNMENTS

RESULT 1  
US-10-716-005-1  
; Sequence 1, Application US/10716005  
; Publication No. US20050106578A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; APPLICANT: Cockerill III, Franklin R.  
; APPLICANT: Aichinger, Christian  
; APPLICANT: Reiser, Astrid  
; TITLE OF INVENTION: Detection of Group B Streptococcus  
; FILE REFERENCE: 070319/460001  
; CURRENT APPLICATION NUMBER: US/10/716,005  
; CURRENT FILING DATE: 2003-11-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 23  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-716-005-1

Query Match 100.0%; Score 23; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23

Db 1 TGAGAGGCGAGTAGAAGCTTAG 23

```
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 4
US-10-465-205-13

Query Match      86.1%; Score 19.8; DB 7; Length 1379;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1 TGAGAAGGCAGTAGAAGCTTAG 23
     ||| ||| ||| ||| ||| ||| |||
Db    9 TGAAAATGTCAGTAGAAGCTTAG 31

RESULT 5
US-10-465-205-16
; Sequence 16, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 1
US-10-465-205-16

Query Match      86.1%; Score 19.8; DB 7; Length 1379;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1 TGAGAAGGCAGTAGAAGCTTAG 23
     ||| ||| ||| ||| ||| ||| |||
Db    10 TGAAAATGTCAGTAGAAGCTTAG 32

RESULT 6
US-10-465-205-8
; Sequence 8, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
```

```
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 5
US-10-465-205-8

Query Match      86.1%; Score 19.8; DB 7; Length 1384;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAAAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 7
US-10-465-205-10
; Sequence 10, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 8
US-10-465-205-10

Query Match      86.1%; Score 19.8; DB 7; Length 1384;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 9 TGAAAATGCAGTAGAAGCTTAG 31
    |||||

RESULT 8
US-10-465-205-9
; Sequence 9, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 7
US-10-465-205-9

Query Match      86.1%; Score 19.8; DB 7; Length 1385;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAAAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 9
US-10-465-205-11
; Sequence 11, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 9
US-10-465-205-11

Query Match      86.1%; Score 19.8; DB 7; Length 1390;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 24 TGAAAATGCAGTAGAAGCTTAG 46
    |||||

RESULT 10
US-10-465-205-15
; Sequence 15, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 3
US-10-465-205-15

Query Match      86.1%; Score 19.8; DB 7; Length 1393;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAAAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 11
US-10-465-205-17
; Sequence 17, Application US/10465205
```

```

; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURES:
; OTHER INFORMATION: ptSI sequence from isolate no. 11
US-10-465-205-17

Query Match      86.1%; Score 19.8; DB 7; Length 1447;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGCGCAGTAGAAAGCTTAG 32
   ||| ||| ||| ||| ||| ||| ||| |||
Db 10 TGAATATGCAGTAGAAAGCTTAG 32

RESULT 12
US-10-282-122A-38442
; Sequence 38442, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

```



```
Best Local Similarity 90.9%; Pred. No. 1.5e+02; Indels 2; Mismatches 0; Gaps 0;
Matches 20; Conservative 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
|||||
Db 525 GAGAAGGAGTAGAAGCCTAG 504

RESULT 15
US-10-741-600-17691/c
; Sequence 17691, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michelle et al.
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17691
; LENGTH: 81210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17691

Query Match 81.7%; Score 18.8; DB 8; Length 81210;
Best Local Similarity 90.9%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
|||||
Db 525 GAGAAGGAGTAGAAGCCTAG 504

RESULT 16
US-10-504-582-15/c
; Sequence 15, Application US/10504582
; Publication No. US20050176943A1
; GENERAL INFORMATION:
; APPLICANT: Yoshitake NISHIMUNE
; APPLICANT: Hiromitsu TANAKA
; APPLICANT: Masami NOZAKI
; TITLE OF INVENTION: Mouse spermatogenesis genes, mutations of male infertility-related
; TITLE OF INVENTION: and uses thereof.
; FILE REFERENCE: 2004-1256A/WMC/00653
; CURRENT APPLICATION NUMBER: US/10/504,582
; CURRENT FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: JP2002-36649
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: JP2002-381241
; PRIOR FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-504-582-15

Query Match 79.1%; Score 18.2; DB 9; Length 720;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
Db 499 TGAGAAGGCAGTAGAAGCAGAG 477

RESULT 17
US-10-473-939-5
; Sequence 5, Application US/10473939
```

```
Publication No. US20040235717A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Chawnsiang
; TITLE OF INVENTION: SUPPRESSION OF ANDROGEN RECEPTOR
; TITLE OF INVENTION: TRANSACTIVATION THROUGH NEW PATHWAYS TO AR AND AR
; FILE REFERENCE: 21108.001304
; CURRENT APPLICATION NUMBER: US/10/473,939
; CURRENT FILING DATE: 2003-10-03
; PRIOR APPLICATION NUMBER: PCT/US02/11086
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 60/365,060
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 60/282,265
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 160820
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note =
; OTHER INFORMATION: synthetic construct
US-10-473-939-5
```

```
Query Match 79.1%; Score 18.2; DB 8; Length 160820;
Best Local Similarity 87.0%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
Db 55080 TGAGAAGGAGTAGGAAGCTTTG 55102
```

```
RESULT 18
US-09-925-065A-729139/c
; Sequence 729139, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 729139
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-729139
```

```
Query Match 77.4%; Score 17.8; DB 4; Length 522;
Best Local Similarity 90.5%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
|||||
Db 425 TGAGAAGGCAGAGAGAGCTT 405
```

RESULT 19  
US-09-925-065A-739199/c  
; Sequence 739199, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 739199  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-739199

Query Match 77.4%; Score 17.8; DB 4; Length 525;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

DB 425 TGAGAAGGCAGTAGAAGCTT 405

RESULT 20  
US-09-925-065A-728631/c  
; Sequence 728631, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 728631  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-728631

Query Match 77.4%; Score 17.8; DB 4; Length 545;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

DB 496 TGAGAAGGCAGTGGGAAGCTT 476  
RESULT 21  
US-09-925-065A-69405/c  
; Sequence 69405, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69405  
; LENGTH: 624  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-69405

Query Match 77.4%; Score 17.8; DB 4; Length 624;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

DB 424 TGAGAAGGCAGTAGAAGCTT 404

RESULT 22  
US-09-925-065A-8982/c  
; Sequence 8982, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8982  
; LENGTH: 863  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-8982

Query Match 77.4%; Score 17.8; DB 4; Length 863;  
Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 469 TGAGAAGGCAGTGGGAAGCTT 449

RESULT 23  
US-09-925-065A-709349/c  
; Sequence 709349, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 709349  
; LENGTH: 1206  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-709349

Query Match 77.4%; Score 17.8; DB 4; Length 1206;  
Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 752 TGAGAAGGCAGTAGAAGCTT 732

RESULT 24  
US-09-925-065A-709350/c  
; Sequence 709350, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 709350  
; LENGTH: 1206  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-709350

Query Match 77.4%; Score 17.8; DB 4; Length 1206;  
Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 752 TGAGAAGGCAGTAGAAGCTT 732

RESULT 25  
US-09-925-065A-62246  
; Sequence 62246, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 62246  
; LENGTH: 1253  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-62246

Query Match 77.4%; Score 17.8; DB 4; Length 1253;  
Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 109 TGAGAAGGCAGTGGGAAGCTT 129

RESULT 26  
US-10-425-114-24004  
; Sequence 24004, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 24004  
; LENGTH: 1714  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3606-055-G8\_FLI  
US-10-425-114-24004

Query Match 77.4%; Score 17.8; DB 7; Length 1714;

Best Local Similarity 90.5%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;  
Matches 19; Conservative 0;  
QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 968 TGAGAAGGCAGTAGAAGCTT 988

RESULT 27  
US-10-027-632-97987/c  
; Sequence 97987, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 97987  
; LENGTH: 1940  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-97987

Query Match 77.4%; Score 17.8; DB 5; Length 1940;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 469 TGAGAAGGCAGTAGAAGCTT 449

RESULT 28  
US-10-027-632-97987/c  
; Sequence 97987, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 97987  
; LENGTH: 1940  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-97987

Query Match 77.4%; Score 17.8; DB 6; Length 1940;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 469 TGAGAAGGCAGTAGAAGCTT 449

RESULT 29  
US-10-425-115-112272  
; Sequence 112272, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 112272  
; LENGTH: 1991  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_33880C.1  
US-10-425-115-112272

Query Match 77.4%; Score 17.8; DB 8; Length 1991;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 969 TGAGAAGGCAGTAGAAGCTT 989

RESULT 30  
US-10-437-963-95439  
; Sequence 95439, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 95439  
; LENGTH: 2478  
; TYPE: DNA

```

; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MKT4530_93633C.1
US-10-437-963-95439

Query Match          77.4%; Score 17.8; DB 7; Length 2478;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 1638 TGAGAAGGCAGTAGAAGCTT 1658

RESULT 31
US-10-108-260A-1480/c
; Sequence 1480, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1480
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1480

Query Match          77.4%; Score 17.8; DB 6; Length 2533;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 702 TGAGAAGGCAGTAGAAGCTT 682

RESULT 32
US-10-027-632-114128/c
; Sequence 114128, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 114128
; LENGTH: 3279
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114128

Query Match          77.4%; Score 17.8; DB 5; Length 3279;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 424 TGAGAAGGCAGTAGAAGCTT 404

RESULT 34
US-10-027-632-114128/c
; Sequence 114128, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 114128
; LENGTH: 3279
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114128
```

; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114128  
; LENGTH: 3279  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-114128

Query Match 77.4%; Score 17.8; DB 6; Length 3279;  
Best Local Similarity 90.5%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 424 TGAGAAGGCAGTAGAGAGCTT 404

RESULT 35  
US-10-027-632-114129/c  
; Sequence 114129, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108927.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114129  
; LENGTH: 3279  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-114129

Query Match 77.4%; Score 17.8; DB 6; Length 3279;  
Best Local Similarity 90.5%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 424 TGAGAAGGCAGTAGAGAGCTT 404

RESULT 36  
US-11-097-143-16906  
; Sequence 16906, Application US/11097143  
; Publication No. US2005020858A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.

; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16906  
; LENGTH: 6805  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-16906

Query Match 77.4%; Score 17.8; DB 10; Length 6805;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 2650 TGAGAAGGCAGTAGAAGCTT 2670

RESULT 37  
US-10-741-601-5754/c  
; Sequence 5754, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001500  
; CURRENT APPLICATION NUMBER: US/10/741,601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5754  
; LENGTH: 17391  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-741-601-5754

Query Match 77.4%; Score 17.8; DB 7; Length 17391;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 9559 TGAGAAGGCAGTAGAGAGCTT 9539

RESULT 38  
US-10-741-601-5677  
; Sequence 5677, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001500

```
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5677
; LENGTH: 48265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(48265)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5677

Query Match          77.4%; Score 17.8; DB 7; Length 48265;
Best Local Similarity 90.5%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGAGAGGCGAGTAGAAGCTT 21
      |||||
Db      2089 TGAGAGGCGAGTAGAAGCTT 2109

RESULT 39
US-10-322-281-123/c
; Sequence 123, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123
; LENGTH: 104399
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(104399)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-123

Query Match          77.4%; Score 17.8; DB 7; Length 104399;
Best Local Similarity 90.5%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GAGAAGGCGAGTAGAAGCTTA 22
      |||||
Db      35008 GAGAAGGCGAGTAGAAGACTA 34988

RESULT 40
US-10-465-205-1
; Sequence 1, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-10-465-205-1

Query Match          75.7%; Score 17.4; DB 7; Length 21;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 AAGGCGAGTAGAAGCTTAG 23
      |||||
Db      2 AATGCGAGTAGAAGCTTAG 20

RESULT 41
US-10-469-285-225/c
; Sequence 225, Application US/10469285
; Publication No. US20040126776A1
; GENERAL INFORMATION:
; APPLICANT: LION Bioscience AG
; TITLE OF INVENTION: Gene library
; FILE REFERENCE: L 1541
; CURRENT APPLICATION NUMBER: US/10/469,285
; CURRENT FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 840
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Mus Musculus
; ORGANISM: Mus Musculus
US-10-469-285-225

Query Match          75.7%; Score 17.4; DB 7; Length 309;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGAGAAGGCGAGTAGAAGC 19
      |||||
Db      252 TGAGAAGGCGAGCAGAAAGC 234

RESULT 42
US-09-917-800A-1065/c
; Sequence 1065, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
```

; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1065  
; LENGTH: 322  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1179539  
US-09-917-800A-1065

Query Match 75.7%; Score 17.4; DB 3; Length 322;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGC 19  
|||||  
DB 252 TGAGAAGGCAGCAGAAAGC 234

RESULT 43  
US-09-925-065A-560212/c  
; Sequence 560212, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 560212  
; LENGTH: 569  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-560212

Query Match 75.7%; Score 17.4; DB 4; Length 569;  
Best Local Similarity 94.7%; Pred. No. 3.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTT 21  
|||||  
DB 345 AAAAGGCAGTAGAAGCTT 327

RESULT 44  
US-10-764-420-2470  
; Sequence 2470, Application US/10764420  
; Publication No. US20050084872A1  
; GENERAL INFORMATION:  
; APPLICANT: Lum, Pek Yee  
; APPLICANT: Tan, Yejun  
; APPLICANT: Dai, Hongyue  
; TITLE OF INVENTION: Methods For Determining Whether An Agent  
; FILE OF INVENTION: Possesses A Defined Biological Activity  
; FILE REFERENCE: ROSA122057  
; CURRENT APPLICATION NUMBER: US/10/764,420  
; CURRENT FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US 60/442,797  
; PRIOR FILING DATE: 2003-01-24

; PRIOR APPLICATION NUMBER: US 60/474,413  
; PRIOR FILING DATE: 2003-05-30  
; NUMBER OF SEQ ID NOS: 3683  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2470  
; LENGTH: 2151  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-764-420-2470

Query Match 75.7%; Score 17.4; DB 9; Length 2151;  
Best Local Similarity 94.7%; Pred. No. 4.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGC 19  
|||||  
DB 1891 TGAGAAGGCAGCAGAAAGC 1909

RESULT 45  
US-11-097-143-40201  
; Sequence 40201, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; FILE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 40201  
; LENGTH: 24789  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-40201

Query Match 75.7%; Score 17.4; DB 10; Length 24789;  
Best Local Similarity 94.7%; Pred. No. 5.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGC 19  
|||||  
DB 13733 TGAGAAGGCAGCAGAAAGC 13751

RESULT 46  
US-10-535A-45247  
; Sequence 45247, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis



```

; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIORITY APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIORITY APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIORITY APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIORITY APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45247
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-45247

```

Query Match	74.8%	Score 17.2;	DB 7;	Length 300;
Best Local Similarity	86.4%	Pred. No. 4.4e+02;		
Matches 19:	Conservative	0;	Mismatches 3;	Indels 0; Gaps 0;

QY 2 GAGAAAGCAGTAGAGAAAGCTTAG 23  
137 GAGGAGGGAGTAGAGAAAGCTGAG 158  
DB

```

RESULT 47
US/10-085-783A-45247
; Sequence 45247, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent version 3.2
; SEQ ID NO 45247
; LENGTH: 300

```

;  
; TYPE: DNA  
;  
; ORGANISM: Human  
IIS-10-085-783A-45247

Query Match :	74.8%	Score 17.2;	DB 7;	Length 300;
Best Local Similarity	86.4%;	Pred. No. 4.4e+02;		
Matches 19:	Conservative	0;	Mismatches 3;	Indels 0;
	Gaps	0;		

Qy 2 GAGAAGGCAGTAGAAAAGCTTAG 23  
||| ||| ||| ||| ||| ||| |||  
Db 137 GAGGAGGGAGTAGAAAAGCTGAG 158

RESULT 48  
US-10-062-254-187/c  
; Sequence 187, Application US/10062254  
; Publication No. US2002013882A1  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Edgar B  
; APPLICANT: Cahoon, Rebecca E  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Fang, Yiwen  
; APPLICANT: Hantke, Sabine S.  
; APPLICANT: Lee, Jian-Ming

/ APPLICANT: Li, Zhongsen  
 / APPLICANT: Miao, Guo-Hua  
 / APPLICANT: Morgante, Michele  
 / APPLICANT: Niu, Xiping  
 / APPLICANT: Odell, Joan  
 / APPLICANT: Rafalski, Antoni  
 / APPLICANT: Sakai, Hajime  
 / APPLICANT: Zheng, Peizhong  
 / APPLICANT: Zhu, Qun  
 / TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved In Plant Metabolism  
 / FILE REFERENCE:  
 / CURRENT APPLICATION NUMBER: US/10/062,254  
 / CURRENT FILING DATE: 2002-02-01  
 / PRIOR APPLICATION NUMBER: 09/630,346  
 / PRIOR FILING DATE: 2000-07-28  
 / PRIOR APPLICATION NUMBER: 60/146511  
 / PRIOR FILING DATE: 1999-07-30  
 / PRIOR APPLICATION NUMBER: 60/156006  
 / PRIOR FILING DATE: 1999-09-23  
 / PRIOR APPLICATION NUMBER: 60/156899  
 / PRIOR FILING DATE: 1999-09-30  
 / PRIOR APPLICATION NUMBER: 60/157287  
 / PRIOR FILING DATE: 1999-10-01  
 / PRIOR APPLICATION NUMBER: 60/169767  
 / PRIOR FILING DATE: 1999-12-09  
 / PRIOR APPLICATION NUMBER: 60/171054  
 / PRIOR FILING DATE: 1999-12-16  
 / PRIOR APPLICATION NUMBER: 60/172958  
 / PRIOR FILING DATE: 1999-12-21  
 / PRIOR APPLICATION NUMBER: 60/171515  
 / PRIOR FILING DATE: 1999-12-22  
 / PRIOR APPLICATION NUMBER: 60/173535  
 / PRIOR FILING DATE: 1999-12-29  
 / NUMBER OF SEQ ID NOS: 375  
 / SOFTWARE: Microsoft Office 97  
 / SEQ ID NO 187  
 / LENGTH: 494  
 / TYPE: DNA  
 / ORGANISM: Glycine max  
 / FEATURE:  
 / NAME/KEY: unsure  
 / LOCATION: (392)  
 / NAME/KEY: unsure  
 / LOCATION: (409)  
 / NAME/KEY: unsure  
 / LOCATION: (449)  
 / NAME/KEY: unsure  
 / LOCATION: (464)  
 / US-10-062-254-187

Query Match 74.8%; Score 17.2; DB 5; Length 494;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19: Conservative 0; Mismatches 3; Indels

Qy 1 TGAGAAGGCAGTAGAAAGCTTA 22  
|||||  
db 32 TGAGAAGGCAGTAGGAAACTTA 11

RESULT 49  
US-09-917-800A-436  
; Sequence 436, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Archur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicol  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917

```

; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 436
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA965190
; NAME/KEY: misc.feature
; LOCATION: (1)..(519)
; OTHER INFORMATION: n = a or c or g or t
;
US-09-917-800A-436

Query Match          74.8%; Score 17.2; DB 3; Length 519;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2  GAGAAAGCAGTAGAAGCTTAG 23
Db      179  GGGAAAGCAGTAAAGCTTAG 200

RESULT 50
US-10-027-632-237122
; Sequence 237122, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
;
US-10-027-632-237122

Query Match          74.8%; Score 17.2; DB 6; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTAGAAGCTTAA 22
Db      298  TGAGAAGGCAGTAGAATATAA 319

RESULT 52
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
;
US-10-027-632-237122
```

```

Query Match          74.8%; Score 17.2; DB 5; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTAGAAGCTTAA 22
Db      298  TGAGAAGGCAGTAGAATATAA 319

RESULT 51
US-10-027-632-237122
; Sequence 237122, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
;
US-10-027-632-237122

Query Match          74.8%; Score 17.2; DB 6; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCAGTAGAAGCTTAA 22
Db      298  TGAGAAGGCAGTAGAATATAA 319

RESULT 52
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49562
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49562
```

```
Query Match 74.8%; Score 17.2; DB 5; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 GAGAAGGCAGTAGAAGCTTAG 23
||||| ||||| ||||| ||||| |||||
DB 92 GAGAAGGAGGAGGAAGCTCAG 113
```

```
RESULT 53
US-10-027-632-49563
; Sequence 49563, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49563
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49563
```

```
Query Match 74.8%; Score 17.2; DB 5; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 GAGAAGGCAGTAGAAGCTTAG 23
||||| ||||| ||||| ||||| |||||
DB 92 GAGAAGGAGGAGGAAGCTCAG 113
```

```
RESULT 54
US-10-027-632-49564
; Sequence 49564, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49564
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49564
```

```
Query Match 74.8%; Score 17.2; DB 5; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 GAGAAGGCAGTAGAAGCTTAG 23
||||| ||||| ||||| ||||| |||||
DB 92 GAGAAGGAGGAGGAAGCTCAG 113
```

```
RESULT 55
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49562
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49562
```

```
Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 GAGAAGGCAGTAGAAGCTTAG 23
||||| ||||| ||||| ||||| |||||
DB 92 GAGAAGGAGGAGGAAGCTCAG 113
```

```
RESULT 56
US-10-027-632-49563
; Sequence 49563, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49563
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49563

Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAAGCTTAG 23
DB 92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 57
US-10-027-632-49564
; Sequence 49564, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49564
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49564

Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAAGCTTAG 23
DB 92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 58
US-09-925-065A-34
; Sequence 34, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-34

Query Match 74.8%; Score 17.2; DB 4; Length 583;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCGAGTAGAAAGCTTAG 23
DB 249 GAAAAGGCATTAGAAAGCATAG 270

RESULT 59
US-10-972-079-87189
; Sequence 87189, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; CATTLE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 87189
; LENGTH: 600
; TYPE: DNA
```

```
; ORGANISM: Human
US-10-027-632-49564

Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCGAGTAGAAAGCTTAG 23
DB 92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 58
US-09-925-065A-34
; Sequence 34, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-34

Query Match 74.8%; Score 17.2; DB 4; Length 583;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCGAGTAGAAAGCTTAG 23
DB 249 GAAAAGGCATTAGAAAGCATAG 270

RESULT 59
US-10-972-079-87189
; Sequence 87189, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; CATTLE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 87189
; LENGTH: 600
; TYPE: DNA
```

```
; ORGANISM: Chicken 19866894388660_1
US-10-972-079-87189

Query Match          74.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAGCTTAG 23
Db      469 GAAAGGCGAGCAGAAAGCTGAG 490

RESULT 60
US-10-972-079-87190
; Sequence 87190, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 87190
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894388660_2
US-10-972-079-87190

Query Match          74.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAGCTTAG 23
Db      266 GAAAGGCGAGCAGAAAGCTGAG 287
```

Search completed: January 28, 2006, 01:45:28  
Job time : 405.011 secs

**This Page Blank (uspto)**

ECOR1 adaptor) (Start Insert). . . . AAAAAAAAAAAAAAAAAAAAA (End Insert) (Start Tag)TGGCA(End Tag) (Start NotI site/Vector)GGGCGCCACCGG. . . The total number of white colony forming units (cfu) in the primary library before amplification was 2.1x10<sup>6</sup> cfu (colony forming units). The background of empty clones was less than 1%. Inserts ranged from 0.5kb to 4 kb, as determined by PCR. Purified plasmid DNA from the primary library was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites flanking the cloned cDNA inserts. The purified PCR products, representing the entire cloned cDNA population, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30C. Unhybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was 5.6x10<sup>6</sup> cfu. Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 130 AGAAGGCTGTAGAATCTTAG 150

## RESULT 41

BE142442  
LOCUS RC2-HT0148-071099-013-h01 HT0148 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000  
DEFINITION BE142442  
ACCESSION BE142442.1 GI:8605163  
VERSION BE142442.1 GI:8605163  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=RC2-HT0148-071099-013-h01&t3=1999-10-07&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 142.  
Location/Qualifiers  
1. 428  
/organism="Homo sapiens"  
/mol\_type="mRNA"

## FEATURES

source  
1. 428  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0148"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 2; Length 428;  
Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 375 AGAAGGCAGTAGACAGCTTAG 395

## RESULT 42

BM304832  
LOCUS SNEST4a32h03.y1 cSn 1 S neuropa invitro merozoite cDNA Sarcocystis  
DEFINITION BM304832  
ACCESSION BM304832.1 GI:18036536  
VERSION BM304832  
KEYWORDS EST.  
SOURCE Sarcocystis neuropa  
ORGANISM Sarcocystis neuropa  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Sarcocystis.

## REFERENCE

AUTHORS Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, F., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennet, J., Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.  
Sarcocystis neuropa EST project  
Unpublished (2000)  
Contact: Daniel K. Howe  
Sarcocystis neuropa EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 347.  
Location/Qualifiers  
1. 442  
/organism="Sarcocystis neuropa"  
/mol\_type="mRNA"  
/strain="Sn3"  
/db\_xref="taxon:42890"  
/dev\_stage="merozoite"  
/lab\_host="DH10B"

## FEATURES

## source

1. 442  
/organism="Sarcocystis neuropa"  
/mol\_type="mRNA"  
/strain="Sn3"  
/db\_xref="taxon:42890"  
/dev\_stage="merozoite"  
/lab\_host="DH10B"  
/clone\_lib="cSn 1 S neuropa invitro merozoite cDNA"  
/note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; The library was constructed by Dan Howe, University of Kentucky. cDNAs were synthesized from poly(A)+ RNA by oligo d(T) priming and directionally cloned into the Uni-ZAP XR lambda vector. The library was mass excised as phagemids and rescued in SOLR cells. The plasmid library was recovered from the SOLR cells and transformed in mass into DH10B cells for sequencing. WARNING: the library contains a small percentage of cDNAs derived from the bovine host cells."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 3; Length 442;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22  
 |||||  
 DB 201 GAGAGGCGAGTAGAGGCTTA 221

RESULT 43  
 BF323664  
 LOCUS  
 DEFINITION SNEST4a18a11.y1 cSn 1 S neuropa invitro merozoite cDNA Sarcocystis  
 neuropa cDNA 5', mRNA sequence.

ACCESSION BF323664  
 VERSION BF323664.1 GI:11273265  
 KEYWORDS EST.  
 SOURCE Sarcocystis neuropa  
 ORGANISM Sarcocystis neuropa

REFERENCE 1 (bases 1 to 454)  
 Sarcocystidae; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystis.

AUTHORS Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M.,  
 Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J.,  
 Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L.,  
 Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J.,  
 Waterston, R. and Wilson, R.

TITLE Sarcocystis neuropa EST project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Daniel K. Howe

Sarcocystis neuropa EST project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 415.

FEATURES  
 source Location/Qualifiers

1..454  
 /organism="Sarcocystis neuropa"  
 /mol\_type="mRNA"  
 /strain="Sn3"  
 /db\_xref="taxon:42890"  
 /dev\_stage="merozoite"  
 /lab\_host="DH10B"  
 /clone\_lib="cSn 1 S neuropa invitro merozoite cDNA"  
 /notes="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 xhoI; The library was constructed by Dan Howe, University  
 of Kentucky. cDNAs were synthesized from poly(A) + RNA  
 by oligo d(T) priming and directionally cloned into the  
 Uni-ZAP XR lambda vector. The library was mass excised  
 as phagemids and rescued in SOLR cells. The plasmid  
 library was recovered from the SOLR cells and transformed  
 in mass into DH10B cells for sequencing. WARNING: the  
 library contains a small percentage of cDNAs derived from  
 the bovine host cells."

ORIGIN

Query Match 77.4%; Score 17.8; DB 2; Length 454;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22  
 |||||  
 DB 196 GAGAGGCGAGTAGAGGCTTA 216

RESULT 44

BM305256  
 LOCUS  
 DEFINITION SNEST4a38c04.y1 cSn 1 S neuropa invitro merozoite cDNA Sarcocystis  
 neuropa cDNA 5', mRNA sequence.

ACCESSION BM305256  
 VERSION BM305256.1 GI:18036960  
 KEYWORDS EST.  
 SOURCE Sarcocystis neuropa  
 ORGANISM Sarcocystis neuropa

REFERENCE 1 (bases 1 to 501)  
 Sarcocystidae; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystis.

AUTHORS Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M.,  
 Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J.,  
 Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L.,  
 Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J.,  
 Waterston, R. and Wilson, R.

TITLE Sarcocystis neuropa EST project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Daniel K. Howe

Sarcocystis neuropa EST project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 425.

FEATURES  
 source Location/Qualifiers

1..501  
 /organism="Sarcocystis neuropa"  
 /mol\_type="mRNA"  
 /strain="Sn3"  
 /db\_xref="taxon:42890"  
 /dev\_stage="merozoite"  
 /lab\_host="DH10B"  
 /clone\_lib="cSn 1 S neuropa invitro merozoite cDNA"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 xhoI; The library was constructed by Dan Howe, University  
 of Kentucky. cDNAs were synthesized from poly(A) + RNA  
 by oligo d(T) priming and directionally cloned into the  
 Uni-ZAP XR lambda vector. The library was mass excised  
 as phagemids and rescued in SOLR cells. The plasmid  
 library was recovered from the SOLR cells and transformed  
 in mass into DH10B cells for sequencing. WARNING: the  
 library contains a small percentage of cDNAs derived from  
 the bovine host cells."

ORIGIN

Query Match 77.4%; Score 17.8; DB 3; Length 501;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22  
 |||||  
 DB 122 GAGAGGCGAGTAGAGGCTTA 142

RESULT 45

AZ851239/c  
 LOCUS  
 DEFINITION 2M0153G22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0153G22 F, genomic survey sequence.

ACCESSION AZ851239  
 VERSION AZ851239.1 GI:13037038  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

Sciuognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 502)
Dunn,D., Aoyagi,A., Barber,M., Becorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0153 row: G column: 22
Seq primer: CGTGTGAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 502.
Location/Qualifiers
1. .502
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0153G22"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: FWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydronamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pW42 [gi|4732114|gb|AF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 502;
Best Local Similarity 90.5%; Pred. No. 3.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAGGCAGTAGAAGCTTA 22
|||||
Db 315 GAGAAGGCAGTAGAAGACTA 295

RESULT 46
AQ437381
LOCUS
DEFINITION
HS 5122 A1.D03.T7A.RP11-11 Human Male BAC Library Homo sapiens
genomic_clone Plate=698 Col=5 Row=G, genomic survey sequence.
ACCESSION
AQ437381
VERSION
AQ437381.1 GI:4548720
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Sciurognathi; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 522)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 698 row: G column: 5
Seq primer: T7
Class: BAC ends
High quality sequence stop: 522.
Location/Qualifiers
1. .522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=698 Col=5 Row=G"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 522;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGCGCAGTAGAAGCTT 21
|||||
Db 393 TGAGAAGCGCAGTAGAGCTT 413

RESULT 47
BM303377
LOCUS
DEFINITION
SNES74a58c03.y1 cSn 1 S neuona invitro merozoite cDNA Sarcocystis
neuona cDNA 5', mRNA sequence.
ACCESSION
BM303377
VERSION
BM303377.1 GI:18035081
KEYWORDS
EST.
SOURCE
Sarcocystis neuona
ORGANISM
Sarcocystis neuona
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Sarcocystis.
1 (bases 1 to 534)
Howe,D.K., Stamper,S., Tang,K., Sibley,L.D., Clifton,S., Marra,M.,
Hillier,L., Pape,D., Martin,J., Wyllie,T., Theising,B., Bowers,T.,
Gibbons,M., Ritter,E., McCann,R., Blistain,A., Bennet,J.,
Schmitt,A., Ronko,I., Tsagareishvili,R., Fedele,M., Belaygorod,L.,
Franklin,C., Carr,L.M., Grow,A., Maguire,L., Wadkins,J., Richey,J.,
Waterston,R. and Wilson,R.
Sarcocystis neuona EST project
Unpublished (2000)
Contact: Daniel K. Howe

TITLE
JOURNAL
COMMENT

```

Sarcocystis neuropa EST project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gbco  
 High quality sequence stop: 420.

## FEATURES

source

```

1. .534
/organism="Sarcocystis neuropa"
/mol_type="mRNA"
/strain="Sn3"
/db_xref="taxon:42890"
/dev_stage="merozoite"
/lab_host="DH10B"
/clone_lib="cSn 1 S neuropa invitro merozoite cDNA"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; The library was constructed by Dan Howe, University
of Kentucky. cDNAs were synthesized from poly(A)+ RNA
by oligo d(T) priming and directionally cloned into the
Uni-ZAP XR lambda vector. The library was mass excised
as phagemids and rescued in SOLR cells. The plasmid
library was recovered from the SOLR cells and transformed
in mass into DH10B cells for sequencing. WARNING: the
library contains a small percentage of cDNAs derived from
the bovine host cells."

```

## ORIGIN

```

Query Match      77.4%; Score 17.8; DB 3; Length 534;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 2 GAGAGGCGAGTAGAAGCTTTA 22

```

|||||
Db 2 GAGAGGCGAGTAGAAGCTTTA 22

```

## RESULT 48

AQ595160/c

LOCUS

```

DEFINITION      HS_5424_B1_G09_T7A RPCI-11 Human Male BAC Library Homo sapiens
Genomic clone Plate=1000 Col=17 Row=N, Genomic survey sequence.

```

ACCESSION

AQ595160

VERSION

GSS.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homnidae; Homo.

1 (bases 1 to 537)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

10449764

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htec.washington.edu  
 Plate: 1000 row: N column: 17  
 Seq primer: 17  
 Class: BAC ends  
 High quality sequence stop: 537.  
 Location/Qualifiers

## FEATURES

source

```

1. 537
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=1000 Col=17 Row=N"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

```

## ORIGIN

```

Query Match      77.4%; Score 17.8; DB 9; Length 537;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGAGAAGCGCAGTAGAAGCTT 21

```

|||||
Db 438 TGTGAAGCAGTAGAAGCTT 418

```

## RESULT 49

CD313569

LOCUS

DEFINITION

```

CD313569      567 bp mRNA linear EST 16-SEP-2003
StrPu621.003933 Sea urchin unfertilised egg cDNA library MPMGP621
Strongylocentrotus purpuratus cDNA clone
MPMGp621P1813;MPI_621_13P18 5', mRNA sequence.

```

ACCESSION

CD313569

VERSION

EST.

KEYWORDS

Strongylocentrotus purpuratus

ORGANISM

Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidea; Euechinoidea; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 567)

Poustka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A.,

Reinhardt,R., Herwig,R., Panopoulou,G. and Lehrach,H.

Generation, annotation, evolutionary analysis, and database

integration of 20,000 unique sea urchin EST clusters

Genome Res. 13 (12), 2736-2746 (2003)

Contact: Poustka AJ

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: poustka@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONF) to reduce sequencing redundancy. According to the ONF

procedure, clones that display the same hybridisation matrix with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per ONF cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well as the coordinates of

the other clones assigned to the same ONF cluster as the clone from

which the above EST is generated is available at the sea urchin

project web site at: http://www.molgen.mpg.de/ag\_seaurchin/. cDNA

clones and filters are distributed via the Resource Center/Primary

Database of the German Human Genome Project (http://www.rzpd.de)

PCR Primers

FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP) 5'-seq

BACKWARD: 5' GCTATTACCGCAGTCGGAAGGGGATGTG 3' (M13FSP) 3'-seq

Seq primer: 5'-CCGTCGCGAATTCGGGT-3' pSport3/86

High quality sequence stop: 567.

```

FEATURES
  source
    Location/Qualifiers
      1..567
        /organism="Strongylocentrotus purpuratus"
        /mol_type="mRNA"
        /db_xref="taxon:7668"
        /clone="MPMGp621P1813.MPI 621 13P18"
        /issue_type="whole unfertilised eggs"
        /dev_stage="embryonic 0hr"
        /lab_host="S.colli, XLI blue"
        /clone_lib="sea urchin unfertilised egg cDNA library
        MPMGP621"
        /note="Vector: pSport1; Site 1: NotI; Site 2: SalI; Random
        primed and directionally cloned in pSport1 vector using a
        NotI (5'-pGACTAGTTCGATCGCGCGCCGCTT)15-3' and a
        SalI 5'-TCGACCCACCGCTCGG-3'adapters (Gibco BRL)"

ORIGIN
  Query Match          77.4%; Score 17.8; DB 6; Length 567;
  Best Local Similarity 90.5%; Pred. No. 3.4e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGCTT 21
    ||||| ||||| ||||| |||||
Db 29 TGAGGAGCGAGTGAAGCTT 49

RESULT 50
CR221772/c
LOCUS
DEFINITION
  CR221772 579 bp DNA linear GSS 06-JUL-2004
  Reverse strand read from insert in 5'pRT insertion targeting and
  chromosome engineering clone MHPN262d10, genomic survey sequence.

ACCESSION
  CR221772.1 GI:50000621
VERSION
  GSS; genome survey sequence; MICR.
KEYWORDS
  Mus musculus (house mouse)
SOURCE
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Murioidea; Muridae; Murinae; Mus.

REFERENCE
  1 (bases 1 to 579)
  Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
  Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
  Rogers,J. and Bradley,A.
  Direct Submission
  Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. http://www.sanger.ac.uk/MICR

FEATURES
  source
    Location/Qualifiers
      1..579
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="MHPN262d10"
        /clone_lib="MHPN"

ORIGIN
  Query Match          77.4%; Score 17.8; DB 11; Length 579;
  Best Local Similarity 90.5%; Pred. No. 3.4e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 210 AGAAGGAAGTGAAGATTAG 190

RESULT 51
CN496517
LOCUS
DEFINITION
  CN496517 Mdfw2022g09.y1 Mdfw Malus x domestica cDNA clone Mdfw2022g09 5',
  mRNA sequence.
ACCESSION
  CN496517
VERSION
  CN496517.1 GI:46599127
KEYWORDS
  EST.
  SOURCE
    Malus x domestica

```

ORGANISM

Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 585)  
Korban,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
Aldwinckle,H., Malnoy,M., Carroll,N., Goldebroough,P., Orvis,K.,  
Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,  
Tsagarisvili,R., Kennedy,S., Waterston,R. and Wilson,R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estowatson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: A. Hernandez / K. Gasic Library sequenced by:  
Washington University Genome Sequencing Center  
WashU EST name: aaf85d05.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers

FEATURES

source

1..585  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdfw2022g09"  
/issue\_type="flower"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdfw"  
/note="Vector: pBluescript II SK (+); Site\_1: NotI;  
Site 2: EcoRI; Total RNA was extracted separately from  
each floral stage (bud, balloon, open and after  
pollination), using the 'pine tree' method. Poly(A)+mRNA  
was isolated twice from total RNA from each stage using  
the Oligotex Direct mRNA kit (Qiagen). mRNA was reverse  
transcribed into double stranded cDNA using a modified  
oligo18(dT) primer with an identifying tag sequence (see  
table below). cDNA's from different stages were pooled in  
equal amounts before adaptor ligation. Tag identification  
when sequencing from 5' end: Stage 1 (bud) insert  
18(A)TCGCA; Stage 2 (balloon) insert 18(A)TCGA; Stage 3  
(open) insert 18(A)TGCCT; Stage 4 (afterpollination)  
insert 18(A)TCGCT. Tag identification when sequencing from  
3' end: Stage 1 (bud) TCCGA18(T) insert; Stage 2 (balloon)  
TCGA18(T) insert; Stage 3 (open) ACGCA18(T) insert; Stage  
4 (afterpollination) ACCGA18(T) insert. Double stranded  
cDNAs were size selected (more than 450 bp), adaptored  
with EcoRI adapters at both ends and then digested with  
NotI. The cDNAs were then directionally cloned into  
EcoRI-NotI digested pBS II SK(+) phagemid  
vector(Stratagene). Identification of adaptors and tags in  
5'-end sequenced clones: (Vector) . . . TAAGCTT(End  
EcoRI adaptor)(Start EcoRI adaptor)GATATCGATTCTTCATTGCTGGG (End  
EcoRI adaptor)(Start insert) . . . AAAAAAAAAAAAAAAAAA(End  
insert) (Start Tag)TCGA(End Tag)(Start  
NotI site/Vector)GCGCGCCGCCCGG. . . The total number of  
white colony forming units (cfu) in the primary library  
before amplification was 1.1x10<sup>6</sup> cfu (colony forming  
units). The background of empty clones was less than 1%.  
Inserts ranged from 0.5kb to 3 kb, as determined by PCR.  
Purified plasmid DNA from the primary library was  
converted to single-stranded circles and used as a template  
for PCR amplification using the T7 and T3 priming sites  
flanking the cloned cDNA inserts. The purified PCR  
products, representing the entire cloned cDNA population,  
were used as a driver for normalization. Hybridization  
between the single-stranded library and the PCR products  
was carried out for 44 hours at 30C. Unhybridized

single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was  $9 \times 10^6$  cfu. Background of empty clones was less than 1%."

```

ORIGIN
Query Match          77.4%; Score 17.8; DB 7; Length 585;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AGAAGGCAGTGAAGAGCTTAG 23
||||| ||||| ||||| |||||
Db 462 AGAAGGCTGTAGAAATCTTAG 482

RESULT 52
CB638470
LOCUS          624 bp      mRNA      linear      EST 08-APR-2003
DEFINITION    OSJNEa06M21.f OSJNEa Oryza sativa (japonica cultivar-group) cDNA
ACCESSION    CB638470
VERSION      CB638470
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 (bases 1 to 624)
Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Ming,R.A.,
Soderlund,C. and Wang,G.-B.
Large-scale identification of expressed sequence tags involved in
rice and rice blast fungus interaction
Plant Physiol. 138 (1), 105-115 (2005)
1588683
COMMENT
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 3288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: atc agc ggc cgc gat cc
BACKWARD: aat taa ccc tca cta aag gg
Plate: 06 row: M column: 21
Seq primer: atc agc ggc cgc gat cc.
FEATURES
source
1..624
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEa06M21"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEa"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 6 hrs after inoculation with Rice Blast (Che
85061)"

ORIGIN
Query Match          77.4%; Score 17.8; DB 6; Length 624;
Best Local Similarity 90.5%; Pred. No. 3.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTGAAGAGCTT 21
||||| ||||| ||||| |||||
Db 14 TGAGAAGGCAGTGAAGAGCTT 34

ORIGIN
Query Match          77.4%; Score 17.8; DB 6; Length 630;
Best Local Similarity 90.5%; Pred. No. 3.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CD278427
T143C02371F (FHIG:C) Axenic plate culture Betula pendula cDNA 5',
mRNA sequence.
630 bp      mRNA      linear      EST 01-SEP-2003
CD278427
LOCUS          630 bp      mRNA      linear      EST 01-SEP-2003
DEFINITION    T143C02371F (FHIG:C) Axenic plate culture Betula pendula cDNA 5',
mRNA sequence.
ACCESSION    CD278427
VERSION      CD278427.1 GI:34390473
KEYWORDS
SOURCE
ORGANISM
Betula pendula (European white birch)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids 1; Fagales; Betulaceae; Betula.
REFERENCE
1 (bases 1 to 630)
Johansson,T., Le Quere,A., Ahren,D., Soderstrom,B., Erlandsson,R.,
Lundeberg,J., Uhlen,M. and Tunlid,A.
Transcriptional responses of Paxillus involutus and Betula pendula
during formation of ectomycorrhizal root tissue
Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)
14964534
COMMENT
Contact: Johansson,T.
Fungal-Host Interaction Group (FHIG)
Microbial Ecology, Institution of Ecology
Ecology Building, Lund University, SE-223 62 Lund, Sweden
Tel: +46 46 222 45 49
Fax: +46 46 222 41 58
Email: tomas.johansson@mbioekol.lu.se
PCR Primers
FORWARD: P104 (5'-GGGAGCGCGCCATTGTGTT-3')
BACKWARD: P105 (5'-AGTGAGTCGAAATCGCGCC-3')
Seq primer: P104
High quality sequence stop: 630.
FEATURES
source
1..630
/organism="Betula pendula"
/mol_type="mRNA"
/strains="Skuleskogen, Skogsforsek, Sweden"
/db_xref="taxon:3505"
/tissue_type="Root tissue"
/dev_stage="25 days of growth after transfer"
/lab_host="Escherichia coli BM25.8"
/clone_lib="(FHIG:C) Axenic plate culture"
/notes="Vector: pRiplex2; Site 1: SfiI; Site 2: SfiI; This
EST clone is originating from one of three cDNA libraries,
constructed for transcript profiling of the mycorrhizal
interaction between the basidiomycete Paxillus involutus
and Betula pendula (birch). One library represents the
developed and functional mycorrhizal root tissue
(' (FHIG:A) Ectomycorrhiza plate culture'), a second
library represents axenically grown fungus (' (FHIG:B)
Axenic plate culture') and a third library represents
axenically grown plants (' (FHIG:C) Axenic plate culture').
Libraries were analyzed in parallel and 3555 (FHIG:A),
3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20)
ESTs of >99bp have been deposited. The cDNA libraries were
constructed from total RNA using the SMART cDNA library
Construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)
according to the manufacturer's instructions. Full-length
cDNAs were trimmed by SfiI, fractionated and directionally
ligated into (lambda)Triplex2 arms. The lambda library was
converted to a plasmid library via site-specific
recombination at loxp sites in a Cre+ strain (E. coli
BM25.8). Plasmid clones were randomly collected and
analysed by DNA sequencing using a plasmid-specific
forward primer (P104)."
```

QY 1 TGAGAGGCGAGTAAAGCTT 21  
 |||||  
 Db 485 TGAGAGGCGAGTAAAGCTT 505

# RESULT 54

CD219734/c 645 bp mRNA linear EST 21-MAY-2003  
 LOCUS CCL1\_58\_H08.g1\_A007 Callus culture/cell suspension Sorghum bicolor  
 DEFINITION cDNA clone CCCL\_58\_H08\_A007 5', mRNA sequence.

# ACCESSION

CD219734

# KEYWORDS

CD219734.1 GI:30963168

# SOURCE

Sorghum bicolor (sorghum)

# ORGANISM

Sorghum bicolor

# REFERENCE

1 (bases 1 to 645)

# AUTHORS

Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Rathore, K., Eastman, A. and Pratt, L.H.

# TITLE

An EST database from Sorghum: callus culture and cell suspension

# JOURNAL

Unpublished (2003)

# COMMENT

Other ESTs: CCCL 58 H08.bi A007

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A & M University;

sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences

are presented as their reverse complement and have been trimmed to

exclude polyA.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

# REFERENCE

1 (bases 1 to 648)

# AUTHORS

Balijs, V.S., Nascimento, L.U. and McCombie, W.R.

# TITLE

ESTs from Canis familiaris right cardiac ventricle (dog)

# JOURNAL

Unpublished (2004)

# COMMENT

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mccombie@cshl.org.

# FEATURES

Location/Qualifiers

1..648

/organism="Canis familiaris"

/mol\_type="mRNA"

/db\_xref="taxon:9615"

/sex="Unknown"

/tissue\_type="Cardiac muscle"

/dev\_stage="3 month old normal canine"

/lab\_host="XL10 Gold"

/clone\_lib="Right Cardiac Ventricle (DOGEST6)"

/notes="Organ: Heart; Vector: pBluescript II SK; Site\_1:

EcoRI; Site\_2: XhoI; Library constructed using pBluescript

XR kit from Stratagene. Cloned cDNA was size selected

between 1-3 kb. Tissue supplied by Mark Haskins VMD, PhD,

Pathology and Medical Genetics, School of Veterinary

Medicine, University of Pennsylvania, 3800 Spruce Street,

Philadelphia, PA 19104-6051"

# ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 648;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAAAGCTT 21

|||||

Db 207 TGAGAGGCGAGGAGAAAGATT 227

|||||

RESULT 56

AZ649204

LOCUS

DEFINITION

1M0518J21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0518J21 R, genomic survey sequence.

ACCESSION

AZ649204

VERSION

AZ649204.1 GI:11782641

KEYWORDS

GSS.

SOURCE

Mus musculus (house mouse)

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

# REFERENCE

1 (bases 1 to 665)

# AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Kelly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

# JOURNAL

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

# ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 648;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAAAGCTT 21

|||||

Db 207 TGAGAGGCGAGGAGAAAGATT 227

|||||

RESULT 56

AZ649204

LOCUS

DEFINITION

1M0518J21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0518J21 R, genomic survey sequence.

ACCESSION

AZ649204

VERSION

AZ649204.1 GI:11782641

KEYWORDS

GSS.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

# REFERENCE

1 (bases 1 to 665)

# AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Kelly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

# JOURNAL

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

# ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 648;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAAAGCTT 21

|||||

Db 207 TGAGAGGCGAGGAGAAAGATT 227

|||||

RESULT 56

AZ649204

LOCUS

DEFINITION

1M0518J21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0518J21 R, genomic survey sequence.

ACCESSION

AZ649204

VERSION

AZ649204.1 GI:11782641

KEYWORDS

GSS.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

# REFERENCE

1 (bases 1 to 665)

# AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Kelly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

# JOURNAL

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

# ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 648;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAAAGCTT 21

|||||

Db 207 TGAGAGGCGAGGAGAAAGATT 227

|||||

RESULT 56

AZ649204

LOCUS

DEFINITION

1M0518J21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0518J21 R, genomic survey sequence.

ACCESSION

AZ649204

VERSION

AZ649204.1 GI:11782641

KEYWORDS

GSS.

```

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0518 row: J column: 21
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 665.
Location/Qualifiers
1. .665
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0518J21"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (GI|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 665;
Best Local Similarity 90.5%; Pred. No. 3.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAAGCTTAG 23
Db 41 AGAAGGAAGTAGAAAGATTAG 61

RESULT 57
LOCUS CE179177/c 668 bp DNA linear GSS 25-SEP-2003
DEFINITION tigr-gss-dog-17000326760515 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE179177
VERSION CE179177.1 GI:35327228
KEYWORDS GSS:
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 668)
AUTHORS Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200

Tel: 801-838-0208
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. .668
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 668;
Best Local Similarity 90.5%; Pred. No. 3.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAAGGCAGTAGAAAGCTT 21
Db 266 TGAGGAAGGCAGGAGAGATT 246

RESULT 58
LOCUS AQ541187/c 689 bp DNA linear GSS 19-MAY-1999
DEFINITION RPCI-11-344K16.TJ RPCI-11 Homo sapiens genomic clone
ACCESSION AQ541187
VERSION AQ541187.1 GI:4871643
KEYWORDS GSS:
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 689)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: RPCI-11-344K16.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbs@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .689
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GBB:7631967"
/db_xref="taxon:9606"
/clone="RPCI-11-344K16"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 689;

```

JOURNAL COMMENT

Best Local Similarity 90.5%; Pred. No. 3.5e+03; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22  
||||| | ||||| |||||  
Db 404 GAGAAAGTAGTAGAAGCTTA 384

RESULT 59

DR108351/c  
LOCUS JHU162A06L704 704 bp mRNA linear EST 13-JUN-2005  
DEFINITION Canine cardiovascular system biased cDNA Canis familiaris cdNA, mRNA sequence.

ACCESSION DR108351  
VERSION DR108351.1 GI:67567706  
KEYWORDS EST.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 704)  
AUTHORS DiSilvestre,D., Yung,C., Gao,Z., Farukhi,Y., Winslow,R.L. and Tomaselli,G.F.  
TITLE Canine cardiovascular system biased cDNA sequences  
JOURNAL Unpublished (2005)  
COMMENT Contact: Gordon F. Tomaselli  
Johns Hopkins University  
720 Rutland Avenue/Ross 844, Baltimore, MD 21205, USA  
Tel: 4109552774  
Fax: 4105022096  
Email: gtomasel@jhm.edu.

FEATURES source

1..704  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/strain="Mixed"  
/db\_xref="taxon:9615"  
/sex="Mixed"  
/clone\_lib="Canine cardiovascular system biased cDNA"  
/notes="Organ: Mixed; Vector: PCDNA3.1; Site\_1: EcoRI; Site\_2: XhoI; Adult tissue from eye, lung, aorta, pulmonary artery and brain; neonatal tissue from liver, spleen, thymus, lung, kidney, aorta and brain; 50% is from cardiac tissues."

ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 704;  
Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21  
||||| | ||||| |||||  
Db 136 TGAGAGGCGAGTAGAAGATT 116

RESULT 60

CL583020/c  
LOCUS CL583020 739 bp DNA linear GSS 10-JUN-2004  
DEFINITION OB\_Ba0048023.1 OB\_Ba Oryza brachyantha genomic clone  
OB\_Ba0048023 3', genomic survey sequence.

ACCESSION CL583020  
VERSION CL583020.1 GI:48541769  
KEYWORDS GSS.  
SOURCE Oryza brachyantha  
ORGANISM Oryza brachyantha  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 739)  
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.  
TITLE OMNP Project

Unpublished (2004)  
Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 3259  
Email: rwing@genome.arizona.edu  
PCR Primers  
FORWARD: TAA TAC GAC TCA CTA TAG GG  
BACKWARD: CAC TCA TTA GGC ACC CCA  
Insert Length: 145000 Std Error: 0.00  
Plate: 0048 row: 0 column: 23  
Seq primer: CAC TCA TTA GGC ACC CCA  
Class: BAC ends.  
Location/Qualifiers  
1..739  
/organism="Oryza brachyantha"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4533"  
/clone="OB\_Ba0048023"  
/tissue\_type="leaves"  
/dev\_stage="mature"  
/lab\_host="DH10B"  
/clone\_lib="OB\_Ba"  
/note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

ORIGIN  
Query Match 77.4%; Score 17.8; DB 10; Length 739;  
Best Local Similarity 90.5%; Pred. No. 3.6e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21  
||||| | ||||| |||||  
Db 618 TGAGAGGCGAGTAGAACCAT 598

Search completed: January 28, 2006, 01:14:06  
Job time : 1877.03 secs

***This Page Blank (uspto)***



Result No.	Score	Query Match	Length	DB	ID	Description
1	19.8	86.1	1378	3	US-10-081-923-14	Sequence 14, Appl
2	19.8	86.1	1379	3	US-10-081-923-7	Sequence 7, Appl
3	19.8	86.1	1379	3	US-10-081-923-13	Sequence 13, Appl
4	19.8	86.1	1379	3	US-10-081-923-16	Sequence 16, Appl
5	19.8	86.1	1384	3	US-10-081-923-8	Sequence 8, Appl
6	19.8	86.1	1384	3	US-10-081-923-10	Sequence 10, Appl
7	19.8	86.1	1385	3	US-10-081-923-9	Sequence 9, Appl
8	19.8	86.1	1390	3	US-10-081-923-11	Sequence 11, Appl
9	19.8	86.1	1393	3	US-10-081-923-15	Sequence 15, Appl
10	19.8	86.1	1447	3	US-10-081-923-17	Sequence 17, Appl
11	19.8	86.1	1803	3	US-10-081-923-5	Sequence 5, Appl
12	18.2	79.1	85132	3	US-09-949-016-14693	Sequence 14693, A
13	18.2	79.1	119214	3	US-09-949-016-12507	Sequence 12507, A
14	18.2	79.1	1664976	3	US-08-916-421B-1	Sequence 1, Appl
15	18.2	79.1	1664976	3	US-09-692-570-1	Sequence 1, Appl
16	17.8	77.4	222691	3	US-09-949-016-11762	Sequence 11762, A
17	17.8	77.4	222697	3	US-09-949-016-15842	Sequence 15842, A
18	17.8	77.4	1664976	3	US-08-916-421B-1	Sequence 1, Appl
19	17.8	77.4	1664976	3	US-09-692-570-1	Sequence 1, Appl
20	17.4	75.7	21	3	US-10-081-923-1	Sequence 1, Appl
21	17.4	75.7	504	3	US-09-621-976-9879	Sequence 9879, Ap
22	17.2	74.8	601	3	US-09-949-016-153906	Sequence 153906,
23	17.2	74.8	3929	3	US-09-799-451-884	Sequence 884, App
24	17.2	74.8	10411	3	US-08-961-527-89	Sequence 89, Appl

```
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 1
US-10-081-923-16

Query Match      86.1%; Score 19.8; DB 3; Length 1379;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||| ||||| ||||| |||||
DB   10 TGAATGTCAGTACAAAGCTTAG 32

RESULT 5
US-10-081-923-8
; Sequence 8, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 5
US-10-081-923-8

Query Match      86.1%; Score 19.8; DB 3; Length 1384;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||| ||||| ||||| |||||
DB   10 TGAATGTCAGTACAAAGCTTAG 32

RESULT 6
US-10-081-923-10
; Sequence 10, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 8
US-10-081-923-10

Query Match      86.1%; Score 19.8; DB 3; Length 1384;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||| ||||| ||||| |||||
DB   9 TGAATGTCAGTACAAAGCTTAG 31
```

```
RESULT 7
US-10-081-923-9
; Sequence 9, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptSI sequence from isolate no. 7
US-10-081-923-9
Query Match      86.1%; Score 19.8; DB 3; Length 1385;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 8
US-10-081-923-11
; Sequence 11, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptSI sequence from isolate no. 9
US-10-081-923-11
Query Match      86.1%; Score 19.8; DB 3; Length 1390;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 9
US-10-081-923-15
; Sequence 15, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptSI sequence from isolate no. 3
US-10-081-923-15
Query Match      86.1%; Score 19.8; DB 3; Length 1393;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 10
US-10-081-923-17
; Sequence 17, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptSI sequence from isolate no. 11
US-10-081-923-17
Query Match      86.1%; Score 19.8; DB 3; Length 1447;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 11
US-10-081-923-5
; Sequence 5, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptSI sequence from Oklahoma University M1 strain
; PUBLICATION INFORMATION:
; AUTHORS: Ferretti et al.
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 98
; PAGES: 4658-4663
US-10-081-923-5
Query Match      86.1%; Score 19.8; DB 3; Length 1447;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32
```



```

; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-918-421B-1

```

Query Match 79.1%; Score 18.2; DB 3; Length 1664976;

Best Local Similarity 87.0%; Pred. No. 1.4e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 169531 TGAAGGCATTAAAAAGCTTAG 169553

```

# RESULT 15

US-09-692-570-1

; Sequence 1, Application US/09692570

; Patent No. 6797466

; GENERAL INFORMATION:

; APPLICANT: Bult et al.

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus

; Patent No. 6797466

; TITLE OF INVENTION: jannaeschii

; FILE REFERENCE: PB275C1

; CURRENT APPLICATION NUMBER: US/09/692,570

; CURRENT FILING DATE: 2003-01-14

; PRIOR APPLICATION NUMBER: US 60/024,428

; PRIOR FILING DATE: 1996-08-22

; PRIOR APPLICATION NUMBER: US 08/916,421

; PRIOR FILING DATE: 1997-08-22

```
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
;
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
Query Match 79.1%; Score 18.2; DB 3; Length 1664976;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 TGAGAGGCGAGTAGAGGCTTAG 23
Db 169531 TGAGAGGCGATTAGAGGCTTAG 169553
```

```
RESULT 16
US-09-949-016-11762/c
; Sequence 1, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11762
; LENGTH: 222691
; TYPE: DNA
; ORGANISM: Human
```

US-09-949-016-11762

```
Query Match 77.4%; Score 17.8; DB 3; Length 222691;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 GAGAAAGCGAGTAGAGGCTTA 22
Db 38609 GAGAAAGTAGTAGAGGCTTA 38589
```

```
RESULT 17
US-09-949-016-15842/c
; Sequence 15842, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15842
; LENGTH: 222697
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15842
```

```
Query Match 77.4%; Score 17.8; DB 3; Length 222697;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 GAGAAAGCGAGTAGAGGCTTA 22
Db 38614 GAGAAAGTAGTAGAGGCTTA 38594
```

```
RESULT 18
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
```

```
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600592)..(600592)
```

---

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (67435)..(67435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
```



```
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match      77.4%; Score 17.8; DB 3; Length 1664976;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGCGCAGTAGAAAGCTTAG 23
      ||||| ||||| ||||| |||||
Db      1305980 AGAAGCGCAGTAGAAAGCTTAG 1305960

RESULT 19
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; PRIORITY FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIORITY FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIORITY FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 1664976
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
```

Matches	19;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	3	AGAAGGCAGTAGAAAGCTTAG	23						
Db	1305980	AGAAGGCAGATGAAGCTTAG	1305960						
RESULT 20									
US-10-081-923-1									
; Sequence 1, Application US/10081923									
; Patent No. 6593093									
; GENERAL INFORMATION:									
; APPLICANT: Uhl, James R.									
; APPLICANT: Cockerill, Franklin R.									
; TITLE OF INVENTION: Detection of Group A Streptococcus									
; FILE REFERENCE: 07039-306001									
; CURRENT APPLICATION NUMBER: US/10/081,923									
; CURRENT FILING DATE: 2002-07-02									
; NUMBER OF SEQ ID NOS: 17									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 1									
; LENGTH: 21									
; TYPE: DNA									
; ORGANISM: Artificial Sequence									
; FEATURE:									
; OTHER INFORMATION: Oligonucleotide									
US-10-081-923-1									
Query Match 75.7%; Score 17.4; DB 3; Length 21;									
Best Local Similarity 94.7%; Pred. No. 1.4e+00;									
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	5	AAGGCAGTAGAAAGCTTAG	23						
Db	2	AATGCAGTAGAAGCTTAG	20						
RESULT 21									
US-09-621-976-9879									
; Sequence 9879, Application US/09621976									
; Patent No. 6639063									
; GENERAL INFORMATION:									
; APPLICANT: Dumas Milne Edwards, J.B.									
; APPLICANT: Jobert, S.									
; APPLICANT: Giordano, J.V.									
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.									
; FILE REFERENCE: GENSET.054PR2									
; CURRENT APPLICATION NUMBER: US/09/621,976									
; CURRENT FILING DATE: 2000-07-21									
; NUMBER OF SEQ ID NOS: 19335									
; SOFTWARE: Patent.pm									
; SEQ ID NO 9879									
; LENGTH: 504									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-09-621-976-9879									
Query Match 75.7%; Score 17.4; DB 3; Length 504;									
Best Local Similarity 94.7%; Pred. No. 1.8e+00;									
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1	TGAGAAGGCAGTAGAAAGC	19						
Db	10	TGAGAAGGCAGCAGAAAGC	28						
RESULT 22									
US-09-949-016-153906/c									
; Sequence 153906, Application US/09949016									
; Patent No. 6812339									
; GENERAL INFORMATION:									
; APPLICANT: VENTER, J. Craig et al.									
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED									

;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 153906  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-153906

Query Match 74.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
|||||

Db 416 TGAGAAGGCAGTAGAAGCTTA 395  
|||||

RESULT 23

US-09-799-451-884  
;; Sequence 884, Application US/09799451  
;; Patent No. 6783969  
;; GENERAL INFORMATION:  
;; APPLICANT: Tang, Y. Tom  
;; APPLICANT: Zhou, Ping  
;; APPLICANT: Goodrich, Ryle  
;; APPLICANT: Asundi, Vinod  
;; APPLICANT: Ren, Feiyang  
;; APPLICANT: Zhang, Jie  
;; APPLICANT: Xue, Aigong J.  
;; APPLICANT: Zhao, Qing A.  
;; APPLICANT: Wang, Jian-Rui  
;; APPLICANT: Ma, Yunging  
;; APPLICANT: Yamazaki, Victoria  
;; APPLICANT: Chen, Rui-hong  
;; APPLICANT: Wang, Zhiwei  
;; APPLICANT: Wang, Dunrui  
;; APPLICANT: Yang, Yonghong  
;; APPLICANT: Wehrman, Tom  
;; APPLICANT: Ghosh, Reena  
;; APPLICANT: Dmanac, Radoje T.  
;; TITLE OF INVENTION: No. 6783969el Nucleic Acids and  
;; TITLE OF INVENTION: Polypeptides  
;; FILE REFERENCE: 803  
;; CURRENT APPLICATION NUMBER: US/09/799,451  
;; CURRENT FILING DATE: 2001-03-05  
;; NUMBER OF SEQ ID NOS: 948  
;; SOFTWARE: pt\_FL\_genes Version 2.0  
;; SEQ ID NO 884  
;; LENGTH: 3929  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1307)..(1621)  
US-09-799-451-884

Query Match 74.8%; Score 17.2; DB 3; Length 3929;  
Best Local Similarity 86.4%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 3625 GAGAAGGCAGTAGAAGCTTAG 3646  
|||||

RESULT 24  
US-08-961-527-89  
;; Sequence 89, Application US/08961527  
;; Patent No. 6420135  
;; GENERAL INFORMATION:  
;; APPLICANT: Charles Kunsch  
;; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
;; NUMBER OF SEQUENCES: 391  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESS: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
;; COMPUTER: HP Vectra 486/33  
;; OPERATING SYSTEM: MSDOS version 6.2  
;; SOFTWARE: ASCII Text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/961,527  
;; FILING DATE:  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brookes, A. Anders  
;; REGISTRATION NUMBER: 36,373  
;; REFERENCE/DOCKET NUMBER: PB340P1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (301) 309-8504  
;; TELEFAX: (301) 309-8512  
;; INFORMATION FOR SEQ ID NO: 89:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10411 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-08-961-527-89

Query Match 74.8%; Score 17.2; DB 3; Length 10411;  
Best Local Similarity 86.4%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
|||||

Db 9575 TGAGAAGGCAGTAGAAGCTTA 9596  
|||||

RESULT 25

US-09-949-016-12614/c  
;; Sequence 12614, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12614

Query Match 74.8%; Score 17.2; DB 3; Length 10411;  
Best Local Similarity 86.4%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
|||||

Db 9575 TGAGAAGGCAGTAGAAGCTTA 9596  
|||||

RESULT 25

US-09-949-016-12614/c  
;; Sequence 12614, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12614

```
; LENGTH: 248968
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)....(248968)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12614

Query Match          74.8%; Score 17.2; DB 3; Length 248968;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22
   |||||||
DB 116118 TGAGAAGGCAGTAGAAGCTTA 116097

RESULT 26
US-09-949-016-16061/c
; Sequence 16067, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16061
; LENGTH: 250958
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)....(250958)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16061

Query Match          74.8%; Score 17.2; DB 3; Length 250958;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22
   |||||||
DB 166108 TGAGAAGGCAGTAGAAGCTTA 166087

RESULT 27
US-09-949-016-147767
; Sequence 147767, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 147767
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-147767

Query Match          73.0%; Score 16.8; DB 3; Length 601;
Best Local Similarity 81.8%; Pred. No. 3.4e+02;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   |||||||
DB 292 GAGAAGGCAGTAGAAGCTTAG 313

RESULT 28
US-09-328-352-177/c
; Sequence 177, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 177
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-177

Query Match          73.0%; Score 16.8; DB 3; Length 2061;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
   |||||||
DB 522 TGAGAAGTCAGAAGAAAGCT 503

RESULT 29
US-09-949-016-16577
; Sequence 16577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16577
; LENGTH: 10421
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16577

Query Match          73.0%; Score 16.8; DB 3; Length 10421;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      2847 GAAGGCAGTAGGAAGTTTAG 2866

RESULT 30
US-09-949-016-16578
; Sequence 16578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16578
; LENGTH: 10421
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16578

Query Match      73.0%; Score 16.8; DB 3; Length 10421;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      2847 GAAGGCAGTAGGAAGTTTAG 2866

RESULT 31
US-09-949-016-12787
; Sequence 12787, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12787
; LENGTH: 44554
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12787

Query Match      73.0%; Score 16.8; DB 3; Length 44554;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      36981 GAAGGCAGTAGGAAGTTTAG 37000

RESULT 32
US-09-949-016-12043
; Sequence 12043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12043
; LENGTH: 44555
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12043

Query Match      73.0%; Score 16.8; DB 3; Length 44555;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      36981 GAAGGCAGTAGGAAGTTTAG 37000

RESULT 33
US-09-513-999C-31242
; Sequence 31242, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 31242
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 143
; OTHER INFORMATION: w=a or t
US-09-513-999C-31242

Query Match      72.2%; Score 16.6; DB 3; Length 262;
Best Local Similarity 82.6%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TCAGAGGCAGTAGAAGCTTAG 23
Db      164 TCAGAGGCATTAGAAGGCATAG 186

RESULT 34
US-09-621-976-3711
; Sequence 3711, Application US/09621976
```

```
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3711
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 161..313
; NAME/KEY: misc_feature
; LOCATION: 327
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-3711

Query Match          72.2%; Score 16.6; DB 3; Length 461;
Best Local Similarity 82.6%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 118 TGAGAATGCAGCAGAGAAGATAG 140

RESULT 35
US-09-949-016-195720
; Sequence 195720, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195720
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195720

Query Match          72.2%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 325 TAAGAAGGCAGCAGACATCTTAG 347

RESULT 36
US-09-023-655-881/c
; Sequence 881, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
```

```
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 881:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 852 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1122218
US-09-023-655-881

Query Match          72.2%; Score 16.6; DB 3; Length 852;
Best Local Similarity 82.6%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 828 TGGGAGGCAGCAGGATGCTTAG 806

RESULT 37
US-08-826-611-1
; Sequence 1, Application US/08826611
; Patent No. 6031154
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B.
; APPLICANT: Kanayama, Yoshinori
; TITLE OF INVENTION: Fructokinase Genes and Their Use in
; Metabolic Engineering of Fruit Sweetness
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```



; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15590  
; LENGTH: 7786  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15590

Query Match 72.2%; Score 16.6; DB 3; Length 7786;  
Best Local Similarity 82.6%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 4398 TGAGAAGCCTGTAGAAGTTCAG 4420

## RESULT 41

US-09-949-016-17447  
; Sequence 17447, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17447  
; LENGTH: 41815  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17447

Query Match 72.2%; Score 16.6; DB 3; Length 41815;  
Best Local Similarity 82.6%; Pred. No. 5.7e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 11357 TGAGAAGCCTGTAGAAGTTAAG 11379

## RESULT 42

US-09-949-016-17295/c  
; Sequence 17295, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17295  
; LENGTH: 110402  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17295

Query Match 72.2%; Score 16.6; DB 3; Length 110402;  
Best Local Similarity 82.6%; Pred. No. 6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 17415 TAAGAAGGCAGCAGACATCTTAG 17393

## RESULT 43

US-09-949-016-12741/c  
; Sequence 12741, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12741  
; LENGTH: 110403  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12741

Query Match 72.2%; Score 16.6; DB 3; Length 110403;  
Best Local Similarity 82.6%; Pred. No. 6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 17422 TAAGAAGGCAGCAGACATCTTAG 17400

## RESULT 44

US-09-949-016-11800/c  
; Sequence 11800, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11800



; LENGTH: 115508  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11800

Query Match 72.2%; Score 16.6; DB 3; Length 115508;  
Best Local Similarity 82.6%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 1160 TGAGGAGGCAGGAGGAGGCTCAG 1138

## RESULT 45

US-09-949-016-14826/c  
; Sequence 14826, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14826  
; LENGTH: 115508  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14826

Query Match 72.2%; Score 16.6; DB 3; Length 115508;  
Best Local Similarity 82.6%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 1160 TGAGGAGGCAGGAGGAGGCTCAG 1138

## RESULT 46

US-09-949-016-14827/c  
; Sequence 14827, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14827  
; LENGTH: 115508  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14827

Query Match 72.2%; Score 16.6; DB 3; Length 115508;  
Best Local Similarity 82.6%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 1160 TGAGGAGGCAGGAGGAGGCTCAG 1138

## RESULT 47

US-09-949-016-15182  
; Sequence 15182, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15182  
; LENGTH: 62474  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15182

Query Match 71.3%; Score 16.4; DB 3; Length 62474;  
Best Local Similarity 94.4%; Pred. No. 7.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGC 19  
|||||  
DB 3601 GAGAGGCAGAGAGAGC 3618

## RESULT 48

US-09-949-016-15784/c  
; Sequence 15784, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15784  
; LENGTH: 69709  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(69709)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15784

Query Match 71.3%; Score 16.4; DB 3; Length 69709;

Best Local Similarity 94.4%; Pred. No. 7.2e+02; DB 3; Length 121234;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 14376 GAGAAGGCAGTTGAAAGC 14359  
|||||

RESULT 49  
US-09-949-016-13159/c  
; Sequence 13159, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13159  
; LENGTH: 108060  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13159

Query Match 71.3%; Score 16.4; DB 3; Length 108060;  
Best Local Similarity 94.4%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 104460 GAGAAGGCAGAGAAAGC 104443  
|||||

RESULT 50  
US-09-949-016-14142/c  
; Sequence 14142, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14142  
; LENGTH: 121234  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14142

Query Match 71.3%; Score 16.4; DB 3; Length 121234;  
Best Local Similarity 94.4%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 117634 GAGAAGGCAGAGAAAGC 117617  
|||||

RESULT 51  
US-09-949-016-14141/c  
; Sequence 14141, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14141  
; LENGTH: 140315  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14141

Query Match 71.3%; Score 16.4; DB 3; Length 140315;  
Best Local Similarity 94.4%; Pred. No. 7.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 136715 GAGAAGGCAGAGAAAGC 136698  
|||||

RESULT 52  
US-09-949-016-12449/c  
; Sequence 12449, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12449  
; LENGTH: 146039  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(146039)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12449

Query Match 71.3%; Score 16.4; DB 3; Length 146039;  
Best Local Similarity 94.4%; Pred. No. 7.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 42706 GAGAAGGCAGTTGAAAGC 42689  
|||||

```
RESULT 53
US-09-949-016-14577/c
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGC 19
DB 174425 GAGAGGCAGTAGCAAGC 174408

RESULT 54
US-09-949-016-14578/c
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14578
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGC 19
DB 174425 GAGAGGCAGTAGCAAGC 174408

RESULT 55
US-09-949-016-84403
; Sequence 84403, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84403
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-84403

Query Match 70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGCTTA 22
DB 145 GAGAGGAGGAGTAGAGAGCTTA 165

RESULT 56
US-09-949-016-84404
; Sequence 84404, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84404
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-84404

Query Match 70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGCTTA 22
DB 66 GAGAGGAGGAGTAGAGAGCTTA 86

RESULT 57
US-09-949-016-14579/c
; Sequence 14579, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14579
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14579

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGC 19
DB 174425 GAGAGGCAGTAGCAAGC 174408
```

```
RESULT 57
US-09-949-016-165413
; Sequence 165413, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165413
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165413

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAAGCTTA 22
Db 14 GAGAAGGCAGTAGAAAATTTA 34

RESULT 58
US-09-949-016-165414
; Sequence 165414, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165414
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165414

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAAGCTTA 22
Db 14 GAGAAGGCAGTAGAAAATTTA 34

RESULT 59
US-09-949-016-165415
; Sequence 165415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165415

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAAGCTTA 22
Db 14 GAGAAGGCAGTAGAAAATTTA 34

Search completed: January 28, 2006, 01:18:29
Job time : 81.6067 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 1864.03 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-1  
Perfect score: 23  
Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues  
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	81.7	468	9 AZ279226	AZ279226 RPCI-23-1
C 2	18.8	81.7	636	9 AZ642726	AZ642726 1M0505G22
C 3	18.8	81.7	669	3 BP955145	BP955145 BP955145
C 4	18.8	81.7	680	11 CR892303	CR892303 Sus scrofa
C 5	18.8	81.7	999	10 CL032011	CL032011 CH216-33M
C 6	18.4	80.0	542	10 CE365658	CE365658 tigr-gss-
C 7	18.4	80.0	628	6 CD695839	CD695839 EST12362
C 8	18.4	80.0	656	3 BF508824	BF508824 BB170010B
C 9	18.4	80.0	709	10 CE409091	CE409091 tigr-gss-
C 10	18.4	80.0	815	10 AG551894	AG551894 Mus muscu
C 11	18.2	79.1	357	10 BB873514	BB873514 BB873514
C 12	18.2	79.1	377	10 CG682385	CG682385 ZMMBCC014
C 13	18.2	79.1	429	7 CF982181	CF982181 maj179A02
C 14	18.2	79.1	433	7 CK100703	CK100703 C067P49.5
C 15	18.2	79.1	531	9 BH025854	BH025854 RPCI-24-2
C 16	18.2	79.1	544	5 BU814643	BU814643 N031C09 P
C 17	18.2	79.1	567	6 CA929386	CA929386 MTU2CA.P1
C 18	18.2	79.1	571	6 CF105022	CF105022 mai93C06
C 19	18.2	79.1	601	3 BQ322034	BQ322034 PM2-CT026
C 20	18.2	79.1	615	1 AV270535	AV270535 AV270535
C 21	18.2	79.1	627	9 CE110333	CE110333 tigr-gss-
C 22	18.2	79.1	640	10 CE426147	CE426147 tigr-gss-

C 23	18.2	79.1	644	9 BZ984463	BZ984463 PUF1180TD
C 24	18.2	79.1	671	8 CX179363	CX179363 D07_45-16
C 25	18.2	79.1	731	11 SP6106024	SP6106024 Equus cab
C 26	18.2	79.1	732	5 BU946020	AGENCOURT
C 27	18.2	79.1	740	3 BI737502	BI737502 603358184
C 28	18.2	79.1	772	7 CO431201	CO431201 UI-M-HX0-
C 29	18.2	79.1	775	5 BU466365	BU466365 603268708
C 30	18.2	79.1	819	6 CA494633	CA494633 AGENCOURT
C 31	18.2	79.1	882	10 CZ717363	CZ717363 OC_Ba004
C 32	18.2	79.1	886	8 DT058135	DT058135 AGENCOURT
C 33	18.2	79.1	1381	4 BC037104	BC037104 Mus muscu
C 34	18.2	79.1	1647	4 BC048862	BC048862 Mus muscu
C 35	18.2	79.1	6244	4 BC038223	BC038223 Homo sapi
C 36	17.8	77.4	272	9 BH868412	BH868412 hk25C06.x
C 37	17.8	77.4	314	7 CO747995	CO747995 ShESTbaa6
C 38	17.8	77.4	367	7 CV629435	CV629435 M6frt3101
C 39	17.8	77.4	367	7 CV630812	CV630812 M6frt3105
C 40	17.8	77.4	367	7 CV657147	CV657147 M6frt3109
C 41	17.8	77.4	428	2 BE142442	BE142442 RC2-HT014
C 42	17.8	77.4	442	3 BM304832	BM304832 ShEST4a32
C 43	17.8	77.4	454	2 BF323664	BF323664 ShEST4a18
C 44	17.8	77.4	501	3 BM305256	BM305256 ShEST4a38
C 45	17.8	77.4	502	9 AZ851239	AZ851239 2M0151G22
C 46	17.8	77.4	522	9 AQ437381	AQ437381 HS_51222_A
C 47	17.8	77.4	534	3 BM303377	BM303377 ShEST4a58
C 48	17.8	77.4	537	9 AQ595160	AQ595160 HS_5424_B
C 49	17.8	77.4	567	6 CD313569	CD313569 StrPu621
C 50	17.8	77.4	579	11 CR221772	CR221772 Reverse e
C 51	17.8	77.4	585	7 CN496517	CN496517 Mafw2022g
C 52	17.8	77.4	624	6 CB638470	CB638470 OSUNE006M
C 53	17.8	77.4	630	6 CD278427	CD278427 T143C0237
C 54	17.8	77.4	645	6 CD219734	CD219734 CCCI_58_H
C 55	17.8	77.4	648	8 CX000060	CX000060 1v54H07.b
C 56	17.8	77.4	665	9 AZ649204	AZ649204 1M0518J21
C 57	17.8	77.4	668	9 CE179177	CE179177 tigr-gss-
C 58	17.8	77.4	689	9 AQ541187	AQ541187 RPCI-11-3
C 59	17.8	77.4	704	8 DR108351	DR108351 JHU162A06
C 60	17.8	77.4	739	10 CL583020	CL583020 OB_Ba004

ALIGNMENTS

RESULT 1  
AZ279226/c 468 bp DNA linear GSS 26-JUL-2000  
LOCUS RPCI-23-135F5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-135F5,  
DEFINITION genomic survey sequence.

ACCESSION AZ279226 GI:9498128  
VERSION AZ279226  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,  
Akinret,B., Levine,M., McGann,S., Tsengaye,G., Geer,K., Krol,M., de  
Jong,P. and Fraser,C.M.  
Jong P. and Fraser C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomes  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com) . BAC end page:  
 http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 135 row: F column: 5  
 Seq primer: SP6  
 Class: BAC ends.  
 Location/Qualifiers

#### FEATURES

```

1. .468
    Location/Qualifiers
      1. .468
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-135P5"
        /sex="Female"
        /lab_host="DH10B"
        /clone_lib="RPCI-23"
        /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
        EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI Methyase. Size
        selected DNA was cloned into the pBACe3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
  
```

#### ORIGIN

```

Query Match      81.7%; Score 18.8; DB 9; Length 468;
Best Local Similarity 90.9%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAAGGCGAGTAGAAAGCTTA 22
    ||||| ||||| ||||| |||||
DB 249 TGAGAAAGGCGAGTAGAAAGCTTA 228
    ||||| ||||| ||||| |||||
  
```

#### RESULT 2

```

AZ642726/c
LOCUS              636 bp DNA linear GSS 14-DEC-2000
DEFINITION
  1M0505G22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  clone UUGC1M0505G22 R, genomic survey sequence.
  
```

```

ACCESSION      AZ642726
VERSION        AZ642726.1 GI:11769618
KEYWORDS       GSS.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
  
```

```

REFERENCE
AUTHORS        Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
               Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
               Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
               Niederhausern,A. and Wright,D.,Weiss,R.
  
```

```

TITLE          Mouse whole genome scaffolding with paired end reads from 10kb
               plasmid inserts
  
```

```

JOURNAL        Unpublished (2000)
  
```

```

COMMENT        Contact: Robert B. Weiss
               University of Utah Genome Center
  
```

```

               Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
               84112, USA
               Tel: 801 585 5606
               Fax: 801 585 7177
  
```

```

               Email: ddunn@genetics.utah.edu
  
```

```

               Insert Length: 10000 Std Error: 0.00
  
```

```

               Plate: 0505 row: G column: 22
  
```

```

               Seq primer: CACACAGGAACAGCTATGACC
  
```

```

               Class: plasmid ends
  
```

```

               High quality sequence stop: 636.
  
```

#### FEATURES

```

Location/Qualifiers
1. .636
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  
```

```

/clone="UUGC1M0505G22"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
  
```

#### ORIGIN

```

Query Match      81.7%; Score 18.8; DB 9; Length 636;
Best Local Similarity 90.9%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAAGGCGAGTAGAAAGCTTA 22
    ||||| ||||| ||||| |||||
DB 37 TGAGAAAGGCGAGTAGAAAGCTTA 16
    ||||| ||||| ||||| |||||
  
```

#### RESULT 3

```

BP955145
LOCUS              669 bp mRNA linear EST 10-AUG-2005
DEFINITION
  BP955145 Euphorbia tirucalli callus induced from stem Euphorbia
  tirucalli cDNA clone ETC041C09 5', mRNA sequence.
  
```

```

ACCESSION      BP955145
VERSION        BP955145.1 GI:58204501
KEYWORDS       EST.
  
```

```

SOURCE         Euphorbia tirucalli
  
```

```

ORGANISM       Euphorbia tirucalli
  
```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Euphorbiales; Euphorbiaceae;
Euphorbiae; Euphorbia.
  
```

```

1 (bases 1 to 669)
  
```

```

AUTHORS        Kajikawa,M., Yamato,K.T., Kohzu,Y., Sakata,R., Fukuzawa,H.,
               Uchida,H. and Ohyama,K.
  
```

```

TITLE          Expressed sequence tags from callus of Euphorbia tirucalli: A
               resource for genes involved in triterpenoid and sterol biosynthesis
  
```

```

JOURNAL        Plant Biotechnol. 21, 349-353 (2004)
  
```

```

COMMENT        Contact: Katsuyuki T Yamato
  
```

```

               Graduate School of Biostudies, Division of Integrated Life Science
               Kyoto University
  
```

```

               Sakyo-ku, Kitashirakawaoiwake-cho, Kyoto, Kyoto 606-8502, Japan
  
```

```

               Tel: 81-75-753-6453
  
```

```

               Fax: 81-75-753-6127
  
```

```

               Email: kyamato@lif.kyoto-u.ac.jp.
  
```

```

               Location/Qualifiers
  
```

```

1. .669
  
```

```

  /organism="Euphorbia tirucalli"
  /mol_type="mRNA"
  /db_xref="taxon:142860"
  /clone="ETC041C09"
  /tissue_type="callus induced from stem"
  /clone_lib="Euphorbia tirucalli callus induced from stem"
  
```

#### ORIGIN

```

Query Match      81.7%; Score 18.8; DB 3; Length 669;
Best Local Similarity 90.9%; Pred. No. 1.2e+03;
  
```

Matches	20;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	1	TGAGAAGGCAGTAGAAAAGCTTTA	22						
Db	448	TGAGAAGGCAGCAGAAAAGTTTA	469						
RESULT 4									
CR892303/c									
LOCUS				680 bp	DNA	linear		GSS 22-NOV-2004	
DEFINITION				Sus scrofa BES, genomic survey sequence.					
ACCESSION				CR892303					
VERSION				CR892303.1				GI:55970967	
KEYWORDS				GSS; Bac-end sequence BES; Genome Survey Sequence.					
SOURCE				Sus scrofa				(pig)	
ORGANISM				Sus scrofa					
REFERENCE				1 (bases 1 to 680)					
AUTHORS				Rogel-Gallard,C., Bourgeaux,N., Billault,A., Vaiman,M. and Chardon,P.					
TITLE				Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements					
JOURNAL				Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)					
PUBMED				1049899					
AUTHORS				2 (bases 1 to 680)					
REFERENCE				Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J., Rogel-Gallard,C., Roy,A., Schibler,L. and Milan,D.					
TITLE				A physical map of the swine genome					
JOURNAL				Unpublished					
REFERENCE				3 (bases 1 to 680)					
AUTHORS				Genoscope.					
TITLE				Direct Submission					
JOURNAL				Submitted (19-NOV-2004) Genoscope - Centre National de Sequences					
REFERENCE				BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)					
FEATURES				- Web : www.genoscope.cns.fr)					
source				Location/Qualifiers					
				1. 680					
				/organism="Sus scrofa"					
				/mol_type="genomic DNA"					
				/strain="Large White"					
				/db_xref="taxon:9823"					
				/clone="b10179H06"					
				/sex="male"					
				/cell_type="fibroblast"					
				/clone_lib="SAB"					
				/note="Genoscope sequence ID : IH0AAA17AD12RM1"					
ORIGIN									
Query Match				81.7%	Score 18.8;	DB 11;	Length 680;		
Best Local Similarity				90.9%	Pred: No. 1.2e+03;				
Matches	20;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	2	GAGAAGGCAGTAGAAAAGCTTAG	23						
Db	551	GAGCAGGCAGTAGAAAAGCTTGG	530						
RESULT 5									
CL032011/c									
LOCUS				999 bp	DNA	linear		GSS 31-DEC-2003	
DEFINITION				CH216-33M22 Sp6.1 CH216 Xenopus tropicalis genomic clone					
ACCESSION				CH216-33M22					
VERSION				CL032011					
KEYWORDS				GSS.					
SOURCE									

```

ORIGIN
Query Match      80.0%; Score 18.4; DB 10; Length 542;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAAGGCAGTAGAAGCTTAG 23
|||||
Db 311 GAAGGCAGTAGAAGCTCAG 330

RESULT 7
CD695839
LOCUS      628 bp mRNA linear EST 25-JUN-2003
DEFINITION human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION CD695839
VERSION    CD695839.1 GI:32221773
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominoidea; Homo.
REFERENCE 1 (bases 1 to 628)
AUTHORS  Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
          Zeng,Y.-X.
TITLE    Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL  Unpublished (2003)
COMMENT  Contact: Yixin Zeng
          Cancer Center
          Sun Yat-sen University
          651 Dongfeng Road East, Guangzhou 510060, China
          Tel: 86-1380-9770-743
          Fax: 86-20-8775-4506
          Email: yxzeng@zsusm.edu.cn.
FEATURES
          Location/Qualifiers
            1..628
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /tissue_type="normal nasopharynx"
              /clone_lib="human nasopharynx"
              /notes="ESTs generated from a normal nasopharynx cDNA
              library from southern Chinese"

ORIGIN
Query Match      80.0%; Score 18.4; DB 6; Length 628;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
|||||
Db 383 TAAGAAGGCAGTAGAAGCT 402

RESULT 8
BI508824
LOCUS      656 bp mRNA linear EST 08-APR-2002
DEFINITION BBI170010B20G05.5 Bee Brain Normalized/Subtracted Library, BBI7 Apis
          mellifera cDNA clone BBI170010B20G05 5', mRNA sequence.
ACCESSION BI508824
VERSION    BI508824
KEYWORDS  EST.
SOURCE    BI508824.1 GI:15359198
          Apis mellifera (honey bee)
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
          Apoidea; Apis.
          1 (bases 1 to 656)
REFERENCE 1
AUTHORS  Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
          Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.P.
          Annotated expressed sequence tags and cDNA microarrays for studies
          of brain and behavior in the honey bee

ORIGIN
Query Match      80.0%; Score 18.4; DB 3; Length 656;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
|||||
Db 116 TCAGAAGGCAGTAGAAGCT 135

RESULT 9
CE409091/c
LOCUS      709 bp DNA linear GSS 27-SEP-2003
DEFINITION tigr-gss-dog-17000362259182 Dog Library Canis familiaris genomic,
          genomic survey sequence.
ACCESSION CE409091
VERSION    CE409091.1 GI:36663905
KEYWORDS  GSS.
SOURCE    Canis familiaris (dog)
          Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
          1 (bases 1 to 709)
REFERENCE 1
AUTHORS  Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
          The dog genome: survey sequencing and comparative analysis
          Science 301 (5641), 1898-1903 (2003)
          14512627
          Contact: Kirkness EF

```

```

JOURNAL
PUBMED
COMMENT
Genome Res. 12 (4), 555-566 (2002)
11932240
Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR PRIMERS
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTACCTCACTAAG
Plate: BBI170010B20 row: G column: 05
Seq primer: AGCGGATAACAATTTTCACACAGGA
High quality sequence stop: 656.
FEATURES
          Location/Qualifiers
            1..656
              /organism="Apis mellifera"
              /mol_type="mRNA"
              /strain="mixed strains of European bees, predominantly
              A.m. ligustica"
              /db_xref="taxon:7460"
              /clone="BBI170010B20G05"
              /sex="female"
              /tissue_type="brain"
              /dev_stage="adult worker honey bee"
              /lab_host="DH10B"
              /clone_lib="Bee Brain Normalized/Subtracted Library, BBI17"
              /note="Organ: brain; Vector: p7T3-Pac; Site 1: EcoRI;
              Site 2: NotI; This BBI7 cDNA library was generated by
              subtraction of the BBI6 library with 4000 previously
              sequenced clones. The BBI6 library was contributed by the
              Soares laboratory and it was constructed and normalized
              as described by Bonaldo, M.F., Lennon, G. and Soares,
              M.B. (1996), Genome Research 6(9): 791-806. RNA was
              prepared from dissected brains of adult worker bees of
              various ages and various behavioral groups. "
ORIGIN
Query Match      80.0%; Score 18.4; DB 3; Length 656;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
|||||
Db 116 TCAGAAGGCAGTAGAAGCT 135

RESULT 9
CE409091/c
LOCUS      709 bp DNA linear GSS 27-SEP-2003
DEFINITION tigr-gss-dog-17000362259182 Dog Library Canis familiaris genomic,
          genomic survey sequence.
ACCESSION CE409091
VERSION    CE409091.1 GI:36663905
KEYWORDS  GSS.
SOURCE    Canis familiaris (dog)
          Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
          1 (bases 1 to 709)
REFERENCE 1
AUTHORS  Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
          The dog genome: survey sequencing and comparative analysis
          Science 301 (5641), 1898-1903 (2003)
          14512627
          Contact: Kirkness EF

```



The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

## FEATURES

## Location/Qualifiers

1. .709  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="dog Library"  
/notes="Site 1: BstXI; Libraries were prepared from  
peripheral blood"

## ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 709;  
Best Local Similarity 95.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAAGGCAGTAGAAGCTTAG 23

Db 259 GAAGGCAGTAGAAGCTCAG 240

## RESULT 10

AG551894/c  
LOCUS  
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-468G16.TJ, genomic survey  
sequence.

ACCESSION AG551894

VERSION AG551894.1 GI:48312585

KEYWORDS GSS.

SOURCE Mus musculus molossinus (Japanese wild mouse)

ORGANISM

Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Mus.

## REFERENCE

1 Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T.,  
Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and  
Shiroishi, T.

Contribution of Asian mouse subspecies Mus musculus molossinus to  
genomic constitution of strain C57BL/6J, as defined by BAC-end  
sequence-SNP analysis

Genome Res. 14 (12), 2439-2447 (2004)

JOURNAL

PUBMED 15574823

REFERENCE 2 (bases 1 to 815)

Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

AUTHORS

TITLE

JOURNAL

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/ ,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

COMMENT

Tsukuba Institute, Bio Resource Center  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1. .815

/organism="Mus musculus molossinus"

## FEATURES

## source

## FEATURES

## Location/Qualifiers

e mouse tissues.

/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-468G16.TJ"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"

## ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 815;

Best Local Similarity 95.0%; Pred. No. 1.9e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAAAGCTT 21

Db 757 GAGAGGCAGTAGAAAAGCTT 738

## RESULT 11

BB873514/c

LOCUS

DEFINITION

Mus musculus cDNA clone G630046N07 5', mRNA sequence.

ACCESSION

BB873514

VERSION

BB873514.1 GI:17119724

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 357)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,

Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Tanaka, T., Tomaru, A., Toya, T., Watabiki, A., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

```

source
1. 357
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630046N07"
/sex="male"
/tissue_type="intestinal mucosa"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male
intestinal mucosa"

ORIGIN
Query Match 79.1%; Score 18.2; DB 2; Length 357;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
66 TGAGAAGGCAGTAGGAGCTGAG 44

RESULT 12
CG682385
LOCUS
DEFINITION
CG682385
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 377)
Bharti.A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGIR (2003c)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 102.
Location/Qualifiers
1. .377
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBC0143P02"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBC (EcoRI)"
/notes="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN
Query Match 79.1%; Score 18.2; DB 10; Length 377;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
55 TGAGATGGCAGTAGAAGGCTAAG 77

RESULT 13
source
1. 357
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630046N07"
/sex="male"
/tissue_type="intestinal mucosa"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male
intestinal mucosa"

ORIGIN
Query Match 79.1%; Score 18.2; DB 2; Length 357;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
66 TGAGAAGGCAGTAGGAGCTGAG 44

RESULT 12
CG682385
LOCUS
DEFINITION
CG682385
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 377)
Bharti.A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGIR (2003c)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 102.
Location/Qualifiers
1. .377
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBC0143P02"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBC (EcoRI)"
/notes="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN
Query Match 79.1%; Score 18.2; DB 10; Length 377;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
55 TGAGATGGCAGTAGAAGGCTAAG 77

RESULT 13

```

CF982181/c  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

ORIGIN

Query Match 79.1%; Score 18.2; DB 7; Length 429;  
Best Local Similarity 87.0%; Pred. No. 2.1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||  
Db 225 TGAGAAGGCAGTAGGAAGCAGAG 203

RESULT 14  
CK100703  
LOCUS  
DEFINITION

CF982181  
maj79a02.yl McCarrey Eddy 18 day preleptotene spermatocytes Mus  
musculus cDNA clone IMAGE:7032076 5' similar to TR:Q12706 Q12706  
PROBABLE SERINE/THREONINE PROTEIN KINASE PSK1 ; mRNA sequence.

CF982181  
CF982181.1 GI:38515230  
EST.

Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 429)

McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,  
Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,  
Ritter,E., Tagareishvili,R., Ronko,I., Maguire,L., Kennedy,S.,  
Bennett,J., Waterston,R. and Wilson,R.

NIHES Mouse

Unpublished (2002)

Contact: McCarrey/Eddy NIHES Mouse

NIHES Mouse

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed and donated by J. McCarrey, Ph.D. (Southwest  
Foundation for Biomedical Research, Dept. of Genetics) - excision  
done by E.M. Eddy, Ph.D. (National Institutes of Health, National  
Institute of Environmental Health Sciences)

Possible reversed clone: similarity on wrong strand

Seq primer: Primer name ambiguous.

Location/Qualifiers  
1. .429

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:7032076"

/sex="male"

/tissue\_type="18-day preleptotene spermatocytes"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="McCarrey Eddy 18 day preleptotene

spermatocytes"

/note="Organ: testis; Vector: pBluescript SK+  
(Stratagene); Site 1: XhoII; Site 2: EcoRI; cDNA oligo  
dT-primed [5'-(GA)10-ACTAGTCGAGTTT-TTTT-3'] and  
directionally cloned using 5' linkers 5'-AATTGCGCAGCAG-3'  
and 5'-CTCGTGGCG-3'. Size selection of >400bp material  
gives average insert size ranging from 1-2 kb. Library was  
mass excised (from lambda-UniZAP-XR) and resulting  
single-stranded phagemids were prepped and transformed into  
DH10B. Library constructed and donated by J. McCarrey,  
Ph.D. (Southwest Foundation for Biomedical Research, Dept.  
of Genetics); excision done by E.M. Eddy, Ph.D. (National  
Institutes of Health, National Institute of Environmental  
Health Sciences)."

CK100703  
C067P49.5Pr Populus strain T89 leaves Populus tremula x Populus  
tremuloides cDNA clone C067P49 5', mRNA sequence.



```

RESULT 17
CA929386
LOCUS
DEFINITION
ma193c06.y1 McCarrey Eddy round spermatid Mus musculus cDNA clone
IMAGE:6399754 5' similar to TR:Q12706 Q12706 PROBABLE
SERINE/THREONINE PROTEIN KINASE PSK1 ;, mRNA sequence.
ACCESSION
CF105022
VERSION
CF105022.1 GI:33144089
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 571)
AUTHORS
McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,
Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,
Ritter,B., Tsagarisshvili,R., Ronko,I., Maguire,L., Kennedy,S.,
Bennett,J., Waterston,R. and Wilson,R.
TITLE
NIHS Mouse
JOURNAL
Unpublished (2002)
COMMENT
Contact: McCarrey/Eddy NIHS Mouse
NIHS Mouse
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

RESULT 18
CF105022/c
LOCUS
DEFINITION
mai93c06.y1 McCarrey Eddy round spermatid Mus musculus cDNA clone
IMAGE:6399754 5' similar to TR:Q12706 Q12706 PROBABLE
SERINE/THREONINE PROTEIN KINASE PSK1 ;, mRNA sequence.
ACCESSION
CF105022
VERSION
CF105022.1 GI:33144089
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 571)
AUTHORS
McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,
Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,
Ritter,B., Tsagarisshvili,R., Ronko,I., Maguire,L., Kennedy,S.,
Bennett,J., Waterston,R. and Wilson,R.
TITLE
NIHS Mouse
JOURNAL
Unpublished (2002)
COMMENT
Contact: McCarrey/Eddy NIHS Mouse
NIHS Mouse
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

RESULT 19
B0322034/c
LOCUS
DEFINITION
PM2-CT0265-120301-006-f06 CT0265 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BQ322034
VERSION
BQ322034.1 GI:20931101
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE
1 (bases 1 to 601)
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsumura,A., Balg,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED
10737800
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

```

```

Email: est@watson.wustl.edu
Library constructed and donated by J. McCarrey, Ph.D. (Southwest
Foundation for Biomedical Research, Dept. of Genetics) - excision
done by E.M. Eddy, Ph.D. (National Institutes of Health, National
Institute of Environmental Health Sciences).
Possible reversed clone: similarity on wrong strand
MGI:2497197
Seq primer: Primer name ambiguous.
Location/Qualifiers
1. .571
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:6399754"
/sex="male"
/tissue_type="round spermatids, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/clone_lib="McCarrey Eddy round spermatid"
/note="Organ: testis; Vector: pBluescript SK+
(Stratagene); Site 1: XhoII; Site 2: EcoRI; cDNA oligo
dt-primed [5'-(GA)10-ACTGTCGAGTTTTTTTTT-3'] and
directionally cloned using 5' linkers 5'-AATTCGACAG-3'
and 5'-CTCGTCGCG-3'. Size selection of >400bp material
gives average insert size ranging from 1-2 kb. Library was
mass excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and tranformed
into DH10B. Library contains 98.5% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63423."

```

## ORIGIN

```

Query Match      79.1%; Score 18.2; DB 6; Length 571;
Best Local Similarity 87.0%; Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
Db 424 TGAGAAGGCAGTAGAAGCAGAG 402
|||||

```

## FEATURES source



SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE 1 (bases 1 to 627)  
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
 JOURNAL Science 301 (5641), 1898-1903 (2003)  
 PUBMED 14512627

COMMENT Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

FEATURES Location/Qualifiers  
 source 1..627

/organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_libs="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 627;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 132 TGAGAAGGCAGTTGAACGCTTTG 110

RESULT 22  
 CE426147/c  
 LOCUS tigr-gss-dog-17000362829894 Dog Library Canis familiaris genomic,  
 DEFINITION genomic survey sequence.

ACCESSION CE426147.1 GI:36695469  
 VERSION GSS.  
 KEYWORDS Canis familiaris (dog)  
 SOURCE Canis familiaris  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE 1 (bases 1 to 640)  
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
 JOURNAL Science 301 (5641), 1898-1903 (2003)  
 PUBMED 14512627

COMMENT Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

FEATURES Location/Qualifiers  
 source 1..640  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"

/db\_xref="taxon:9615"  
 /clone\_libs="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 640;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 444 TGAGAAGGCAGTATAAACTTGG 422

RESULT 23  
 BZ984463/c

LOCUS PUF1180TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTa350M15,  
 DEFINITION genomic survey sequence.

ACCESSION BZ984463  
 VERSION BZ984463.1 GI:29219281  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 644)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.

TITLE Maize Genomics Consortium  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: PUF1180TB  
 Contact: Cathy Whitelaw  
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.

FEATURES Location/Qualifiers  
 source 1..644

/organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_libs="ZMMBTa350M15"  
 /clone\_lib="ZM\_0.6\_1.0\_KB"  
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 644;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 39 TGAGAAGGCAGTAGAAGCTTGC 17

RESULT 24  
 CX179363/c

LOCUS CX179363  
 DEFINITION D07\_45-16\_07.ab1 leaf inoculated with Marssonina pathogen of Populus  
 euramericana Populus euramericana cDNA, mRNA sequence.

ACCESSION CX179363  
 VERSION CX179363.1 GI:56826787  
 KEYWORDS EST.  
 SOURCE Populus euramericana

CX179363 671 bp mRNA linear EST 28-DEC-2004

```

ORGANISM
Populus euramericana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE
AUTHORS
1 (bases 1 to 671)
Huang,M., Zhang,X., Zhuge,Q., Zhou,Y., Zhou,H., Zhang,L., Guan,Y.,
Zhang,B. and Wang,M.
TITLE
Identification and Expression Analysis of EST-based Genes in the
Leaf of poplar
JOURNAL
Unpublished (2004)
COMMENT
Contact: Minren Huang
Key Laboratory Tree Genetic Engineering of Nanjing Forestry
University
Nanjing Forestry University
Longpan road 9#, Nanjing, Jiangsu, P.R.China
Tel: 086-025-85427412
Fax: 086-025-85427412
Email: mrhuang@njfu.edu.cn
POLYA=Yes.

FEATURES
source
1..671
/organism="Populus euramericana"
/mol_type="mRNA"
/db_xref="taxon:106131"
/clone_lib="leaf inoculated with Marssonina pathogen of
Populus euramericana"
/notes="vector: pBK-CMV, Site_1: EcoRI; Site_2: XhoI"

ORIGIN
Query Match 79.1%; Score 18.2; DB 8; Length 671;
Best Local Similarity 87.0%; Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 425 TGAGAGGCAGCAAAAGCGTAG 403

RESULT 25
SP6106024/c
LOCUS
DEFINITION
Equus caballus GSS, BAC clone CH241-106024, SP6 end sequence,
genomic survey sequence.
ACCESSION
C7008091
VERSION
C7008091.1 GI:68130624
KEYWORDS
GSS; genomic survey sequence.
SOURCE
Equus caballus (horse)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
REFERENCE
AUTHORS
1 (bases 1 to 731)
Leeb,T., Binns,M.M., de Jong,P.J., Berg,C., Conrad,A., Jarek,M.,
Loehner,T.H., Nordsiek,G., Severitt,S., Scharfe,S.,
Schindewolf,C., Schrader,F., Thies,S. and Bloecker,H.
TITLE
Analysis of horse BAC sequences
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 731)
Jarek,M.
TITLE
Direct Submission
AUTHORS
Submitted (10-JUN-2005) of Genome Analysis, German Research Centre
for Biotechnology, Mascheroder Weg 1, Braunschweig D-38124, Germany
JOURNAL
Location/Qualifiers
FEATURES
source
1..731
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/clone="CH241-106024"
/clone_lib="CHORI-241"
/notes="sequenced with SP6 primer"

ORIGIN
Query Match 79.1%; Score 18.2; DB 11; Length 731;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 471 TGAGAGGCAGTAGAAGCAGAG 449

RESULT 27
BI737502
LOCUS
DEFINITION
603358184F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365317 5',
mRNA sequence.
ACCESSION
BI737502
VERSION
BI737502.1 GI:15714515
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

```

```

Matches
20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 360 TGAGAAATGATAGAAAGCTGAG 338

RESULT 26
BU946020/c
LOCUS
DEFINITION
AGENCOURT_10563999 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6740737 5', mRNA sequence.
ACCESSION
BU946020
VERSION
BU946020.1 GI:24134839
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 732)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1C83076 row: a column: 24
High quality sequence stop: 571.
Location/Qualifiers
1..732
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6740737"
/lab_host="DHI0B (T1-phage-resistant)"
/clone_lib="NIH_MGC_169"
/notes="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggcattatggcc); Site 2: SfiI (ggcgctcgcc); CDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGTCGCGCATTCAGCCGGG-3' and
5'-ATTCTAGAGCCGCGCGGACATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 79.1%; Score 18.2; DB 5; Length 732;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 471 TGAGAGGCAGTAGAAGCAGAG 449

RESULT 27
BI737502
LOCUS
DEFINITION
603358184F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365317 5',
mRNA sequence.
ACCESSION
BI737502
VERSION
BI737502.1 GI:15714515
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

```

```
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 740)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution/LLNL at:
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1929 row: h column: 22
High quality sequence stop: 738.
Location/Qualifiers
1. .740
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5365317"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 79.1%; Score 18.2; DB 3; Length 740;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
DB 696 TGAGAAGGCAGTAGAAGCAGAG 718

RESULT 28
CO431201 772 bp mRNA linear EST 06-JUL-2004
LOCUS UI-W-HX0-csd-e-18-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone
DEFINITION IMAGE:30685457 5', mRNA sequence.
ACCESSION CO431201.1 GI:49677495
VERSION 1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 772)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mouseefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers

FEATURES
source
```

```
1. .772
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30685457"
/tissue_type="whole eye"
/dev_stages="newborn (1, 5, 15 days) and embryonic (15, 16,
17, 18 dpc)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HX0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AATAATTACG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH).".

ORIGIN
Query Match 79.1%; Score 18.2; DB 7; Length 772;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
DB 610 TGAGAAGGCAGTAGAAGCAGAG 632

RESULT 29
BU466365 775 bp mRNA linear EST 30-NOV-2002
LOCUS 603268708F1 CSEQRB20 Gallus gallus cDNA clone CHEST24102 5', mRNA
DEFINITION sequence.
ACCESSION BU466365
VERSION BU466365.1 GI:25959942
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 775)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .775
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST24102"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate
cartilage"

FEATURES
source
```



/dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSQRN20"  
 /note="Vector: pBluescript II KS(+); Site\_1: EcoRI;  
 Site\_2: NotI; This normalized library was constructed from  
 1 million independent clones. cDNA synthesis was initiated  
 using an oligo(dT) primer. Using methylated C in the first  
 strand synthesis reaction. Following this first strand  
 reaction, double-stranded cDNA was blunted, ligated to  
 NotI adapters, digested with EcoRI, size-selected, and  
 cloned into the NotI and EcoRI compatible sites of a  
 custom modified MCS of the pBluescript (KS+) vector. The  
 library was normalized in 2 rounds using conditions  
 adapted from Soares et al., PNAS (1994) 91: 9228-9232 and  
 Bonaldo et al., Genome Research 6 (1996): 791, except that  
 a significantly longer reannealing hybridization was  
 used."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 5; Length 775;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 623 TGAGAAGGCACAGAAAGCTCAG 645

## RESULT 30

CA494633/c  
 LOCUS  
 DEFINITION CA494633 819 bp mRNA linear EST 14-NOV-2002  
 AGENCOURT 10816914 NIH\_MGC\_169 Mus musculus cDNA clone  
 IMAGE:6774068 5', mRNA sequence.

ACCESSION CA494633  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM Mus musculus (house mouse)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 819)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: Dr. Jonathan Kuo, NIMH  
 CDNA Library Preparation: Michael Brownstein Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LUCM3097 row: n column: 19  
 High quality sequence stop: 507.

FEATURES  
 source

1..819  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6774068"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH MGC 169"  
 /note="Torgan: Testicles; Vector: pDNR-LIB; Site 1: SfiI  
 (ggccattagcc); Site 2: SfiI (ggcgctggcc); cDNA made  
 by oligo-dT priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGTGATACGACGATGCGCATACGCCCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.

Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 6; Length 819;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 613 TGAGAAGGCAGTAGAAGCAGAG 591

## RESULT 31

CA494633  
 LOCUS  
 DEFINITION CA494633 882 bp DNA linear GSS 25-JUL-2005  
 OC\_Ba0041115.f OC\_Ba Oryza coarctata genomic clone OC\_Ba0041115  
 5', genomic survey sequence.

ACCESSION CA494633  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM Oryza coarctata (Porteresia coarctata)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 882)  
 Kim H., Collura K., Wisotski M., Byrne M., Stum D., Smart D.,  
 Rao K., Luo M., Jetty R., Kudrna D., Muller C., Soderlund C. and  
 Wing R.  
 OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute  
 Unpublished (2005)  
 Contact: Rod A. Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
 Tel: 520 626 9595  
 Fax: 520 621 1259  
 Email: rwing@genome.arizona.edu  
 PCR Primers  
 FORWARD: TAA TAC GAC TCA CTA TAG GG  
 BACKWARD: CAC TCA TTA GGC ACC CCA  
 Plate: 0041 row: I column: 15  
 Seq primer: TAA TAC GAC TCA CTA TAG GG  
 Class: BAC ends.

FEATURES  
 source

1..882  
 Location/Qualifiers  
 /organism="Oryza coarctata"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:77588"  
 /clone="OC\_Ba0041115"  
 /tissue\_type="leaves"  
 /dev\_stage="mature"  
 /lab\_host="DH10B"  
 /clone\_lib="OC\_Ba"  
 /note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 882;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 Db 742 TGAGAAGGCAGTAGAAGCTCAG 764

## RESULT 32

DT058135  
 LOCUS  
 DEFINITION DT058135 886 bp mRNA linear EST 11-AUG-2005  
 AGENCOURT 55952926 NICHG\_XGC\_Fab xenopus laevis cDNA clone  
 IMAGE:8069608 3', mRNA sequence.  
 ACCESSION DT058135  
 VERSION DT058135.1 GI:72357384

```

KEYWORDS
SOURCE  Xenopus laevis (African clawed frog)
ORGANISM

REFERENCE
1 (bases 1 to 886)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Daniela S. Gerhard, Ph.D.
         Office of Cancer Genomics
         National Cancer Institute / NIH
         Bldg. 31 Rm10A07 Bethesda, MD 20892
         Email: cgabbs-r@mail.nih.gov
         Tissue Procurement: Tom Sargent
         cDNA Library Preparation: Express Genomics
         DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
         DNA Sequencing by: Agencourt Bioscience Corporation
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LLNL at:
         http://image.llnl.gov
         Plate: LLAM17342 row: k column: 14
         High quality sequence start: 398
         High quality sequence stop: 755.
         Location/Qualifiers
             1..886
                 /organism="Xenopus laevis"
                 /mol_type="mRNA"
                 /db_xref="taxon:8355"
                 /clone="IMAGE:8069608"
                 /lab_host="DH10B Tona"
                 /clone_lib="NICHG_XGC_Fab"
                 /note="Organ: fat body; Vector: pExpress-1; Site_1: EcoRV;
                 Site_2: NotI; cDNA was primed using oligo-dT primer:
                 5'-pGACTAGTTCTAGATCCGAGCGCGCCCT(T)25-3' and cloned into
                 the EcoRV/NotI sites of pExpress-1. Size-selection 1.2kb
                 resulted in an average insert size of 1.8kb. This is a
                 primary library (normalized library is NICHG_XGC_FabN) and
                 was constructed by Express Genomics (Frederick, MD). Note:
                 this is a (http://xgc.nci.nih.gov/) Xenopus Gene
                 Collection library."

FEATURES
source
Query Match 79.1%; Score 18.2; DB 8; Length 886;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGGCTTAG 23
||||| ||||||| |||||
Db 742 TGAGAGGCGAGTGAAGGCTTAG 764

RESULT 33
BC037104
LOCUS
DEFINITION Mus musculus cDNA sequence BC030499, mRNA (cDNA clone
IMAGE:5365317), with apparent retained intron.
BC037104
BC037104.1 GI:23331120
HTC.
Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
1 (bases 1 to 1381)
AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klauser,R.D., Collins,F.S., Wagner,J., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Buetow,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettner,M., Madan,A., Young,A.C., Shvachenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvachenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallwood,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL
PUBMED
2 (bases 1 to 1381)
REFERENCE
Director MGC Project.
Direct Submission
Submitted (23-AUG-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nigr.nih.gov
Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granitz,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaapi,R.,
Madsen,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Place: 81 Row: 0 Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene prediction
This clone has the following problem: retained intron.

FEATURES
source
Location/Qualifiers
1..1381
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5365317"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Query Match 79.1%; Score 18.2; DB 4; Length 1381;
Best Local Similarity 87.0%; Pred. No. 2.6e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGGCTTAG 23
||||| ||||||| |||||
Db 711 TGAGAGGCGAGTGAAGGCGAG 733

RESULT 34
BC048862
LOCUS
BC048862 1647 bp mRNA linear HTC 03-AUG-2004

```

```
DEFINITION Mus musculus cDNA sequence BC030499, mRNA (cDNA clone
IMAGE:5367114), with apparent retained intron.
ACCESSION BC048862
VERSION BC048862.1 GI:29179567
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
Worley,K.C., Hale,S., Garcia,A.M., Madan,A., Young,A.C., Shevchenko,Y.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL PUBLISHED 2 (bases 1 to 1647)
REFERENCE DIRECTOR MGC Project.
AUTHORS Direct Submission
TITLE Submitted (14-WAR-2003) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaps-rc@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayèle,K., Becketrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 99 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomesScan gene prediction
This clone has the following problem: retained intron.
Location/Qualifiers
1..1647
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5367114"
/tissue_type="Eye, retina, mouse strain C57Bl/6"

FEATURES
source
/clone_lib="NIH MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Query Match 79.1%; Score 18.2; DB 4; Length 1647;
Best Local Similarity 87.0%; Pred. No. 2.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAGGAAGGCAGTAGAAGAGCTTAG 23
|||||
Db 878 TCAGGAAGGCAGTAGGAGCAGAG 900
|||||

RESULT 35
BC038223
LOCUS BC038223
DEFINITION Homo sapiens DEAH (Asp-Glu-Ala-His) box polypeptide 8, mRNA (cDNA
clone IMAGE:4215266), with apparent retained intron.
ACCESSION BC038223
VERSION BC038223.1 GI:23468302
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 6244)
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL PUBLISHED 2 (bases 1 to 6244)
REFERENCE DIRECTOR MGC Project.
AUTHORS Direct Submission
TITLE Submitted (30-SEP-2002) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaps-rc@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayèle,K., Becketrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 99 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomesScan gene prediction
This clone has the following problem: retained intron.
Location/Qualifiers
1..1647
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5367114"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
```

McDowell, J., Pearson, R., Stantripop, S., Thomas, P. J., Touchman, J. W., Tsugeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 64 Row: a Column: 12  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: retained intron.

#### FEATURES

Location/Qualifiers  
1..6244  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4215266"  
/tissue\_type="Brain, anaplastic oligodendroglioma with 1p/19q loss"  
/clone\_lib="NCI CGAP\_Brn67"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 4; Length 6244;  
Best Local Similarity 87.0%; Pred. No. 3.4e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAGAAAGGCAGTAGAAGCTTAG 23  
|||||  
Db 3421 TCAGAAAGGAATAGAAAGCTTAG 3443

#### RESULT 36

BH868412/C

LOCUS

DEFINITION BH868412 272 bp DNA linear GSS 05-AUG-2002

Genomic clone hk25c06 5', genomic survey sequence.

ACCESSION BH868412

VERSION BH868412.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 272)

AUTHORS Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N., Katzenburger, F., King, L., Miller, B., Muller, S., Nascimeto, L., Zutavern, T., McCombie, W.R. and Martienssen, R.A.

TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)

JOURNAL Unpublished (2002)

COMMENT Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: [mccombie@cshl.org](mailto:mccombie@cshl.org)

Plate: hk25 row: c column: 06

Seq primer: -21M13UnivRev

Class: shotgun

High quality sequence stop: 272.

#### FEATURES

Location/Qualifiers

1..272  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="hk25c06"  
/lab\_host="JM107 or DH5a"  
/clone\_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"  
/note="Organ: Immature ears; Site\_1: Xba I; Site\_2: Xba I;

The vector was digested with XbaI and one nucleotide was added by fill in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were (x/y) reads in M13mp19, b/g reads in pUC19. The same ligation was transformed in either JM107 or DH5a. "

#### ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 272;  
Best Local Similarity 90.5%; Pred. No. 3e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAAAGGCAGTAGAAGCTTA 22

Db 133 GAGAAAGGCAGTAGAAGCTTA 113

#### RESULT 37

CO747995

LOCUS

DEFINITION

CO747995

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

HOWE, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J., Schmitt, A., Ronko, I., Tsagarishvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.

Sarcocystis neuropa EST project

Unpublished (2000)

Contact: Daniel K. Howe

Sarcocystis neuropa EST project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

Contact Daniel K. Howe ([dkhowe2@pop.uky.edu](mailto:dkhowe2@pop.uky.edu)) for further information relating to organism, libraries, or clone availability.

Sequenced by Washington University Genome Sequencing Center

Seq primer: -21UPpot

High quality sequence stop: 314.

Location/Qualifiers

1..314

/organism="Sarcocystis neuropa"

/mol\_type="mRNA"

/strain="SN4"

/db\_xref="taxon:42890"

/dev\_stage="merozoite"

/lab\_host="GC10"

/clone\_lib="Sarcocystis neuropa merozoite UK CSN4 1.cDNA library"

/note="Vector: pDNR-LIB; Site\_1: GGCGCCCTCGGC; Site\_2: GGCCATTACGCC; Library constructed by: Daniel K. Howe and Michelle R. Yeagan Total RNA was isolated from culture-derived merozoites of Sarcocystis neuropa strain SN4. cDNA was synthesized using the template-switching and long-distance PCR method (SMART cDNA library construction kit, BD Biosciences). The amplified cDNA fragments were digested with SfiI, size fractionated, and ligated into SfiI-digested pDNR-LIB vector."

#### ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 314;

Best Local Similarity 90.5%; Pred. No. 3.1e+03; Mismatches 2; Indels 0; Gaps 0;

Matches 19; Conservative 0;

QY 2 GAGAGGCAGTAGAAGCTTA 22  
|||||

Db 40 GAGAGGCAGTAGAGGCTTA 60  
|||||

RESULT 38  
CV629435 367 bp mRNA linear EST 25-OCT-2004  
LOCUS Mdfrt310121.y1 Mdfrt Malus x domestica cDNA clone Mdfrt310121 5',  
DEFINITION mRNA sequence.

ACCESSION CV629435  
VERSION CV629435  
KEYWORDS GI:54622299  
SOURCE EST.  
ORGANISM Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Aldwinckle, H., Malnov, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,  
Tsagarashvili, R., Kennedy, S., Waterston, R., and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)  
Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aal71e11.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdfrt310121"  
/tissue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdfrt"  
/notes="Vector: pBluescript II SK (+); Site 1: NotI;  
each\_stage [young fruitlet (<1cm), young fruitlet (1 cm  
dia.), young fruitlet (12cm dia.), maturing fruit I,  
maturing fruit II, mature fruit], using the 'pine tree'  
method. Poly(A)+mRNA was isolated twice from total RNA  
from each stage using the Oligotex Direct mRNA kit  
(Qiagen). mRNA was reverse transcribed into double  
stranded cDNA using a modified oligo18(dT) primer with an  
identifying tag sequence (see table below). cDNA's from  
different stages were pooled in equal amounts before  
adaptor ligation. Tag identification when sequencing from  
5', end: Stage 1 (young fruitlet) insert 18(A)TCGGT; Stage  
2 (young fruitlet 1cm dia) insert 18(A)TCGGT; Stage 3  
(young fruitlet 12cm dia) insert 18(A)TCGGT; Stage 4  
(maturing fruit I) insert 18(A)TCGGA; Stage 5 (maturing  
fruit II) insert 18(A)TCGGA; Stage 6 (mature fruit) insert  
18(A)TCGGT; Tag identification when sequencing from 3',  
end: Stage 1 (young fruitlet) CACGA18(T) insert; Stage 2  
(young fruitlet 1cm dia) CACGA18(T) insert; Stage 3 (young  
fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing  
fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II)

TCGA18(T) insert; Stage 6 (mature fruit) ACCGA18(T)  
insert. Double stranded cDNAs were size selected (more  
than 450 bp), adapted with EcoRI adaptors at both ends  
and then digested with NotI. The cDNAs were then  
directionally cloned into EcoRI-NotI digested pBS II SK(+)   
phagemid vector(Stratagene). Identification of adaptors  
and tags in 5'-end sequenced clones:  
(Vector) . . . TAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCATGTTGGG (End  
EcoRI adaptor)(Start Insert) . . . AAAAAAAAAAAAAAAAAA (End  
Insert) (Start Tag)TCGGA(End Tag) (Start  
NotI site/Vector)GGGCCGCCACCGCG. . . The total number of  
white colony forming units (cfu) in the primary library  
before amplification was 2.1x10<sup>6</sup> cfu (colony forming  
units). The background of empty clones was less than 1%.  
Inserts ranged from 0.5kb to 4 kb, as determined by PCR.  
Purified plasmid DNA from the primary library was  
converted to single-stranded circles and used as a  
template for PCR amplification using the T7 and T3 priming  
sites flanking the cloned cDNA inserts. The purified PCR  
products, representing the entire cloned cDNA population,  
were used as a driver for normalization. Hybridization  
between the single-stranded library and the PCR products  
was carried out for 44 hours at 30C. Unhybridized  
single-stranded DNA circles were separated from hybridized  
DNA rendered partially double-stranded and electroporated  
into DH10B cells to generate the normalized library. The  
total number of clones with insert was 5.6x10<sup>6</sup> cfu.  
Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 130 AGAAGGCTGTAGAAATCTTAG 150  
|||||

## RESULT 39

CV630812 367 bp mRNA linear EST 25-OCT-2004  
LOCUS Mdfrt3105c20.y1 Mdfrt Malus x domestica cDNA clone Mdfrt3105c20 5',  
DEFINITION mRNA sequence.

ACCESSION CV630812  
VERSION CV630812  
KEYWORDS GI:54623676  
SOURCE EST.  
ORGANISM Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Aldwinckle, H., Malnov, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,  
Tsagarashvili, R., Kennedy, S., Waterston, R., and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)  
Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aal88b10.y1  
Seq primer: -40UP from Gibco.

FEATURES  
source

Location/Qualifiers  
1. .367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdfrt3105c20"  
/issue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdfrt"

/note="Vector: pBluescript II SK (+); Site 1: NotI; Site 2: EcoRII; Total RNA was extracted separately from each stage (young fruitlet (<1cm), young fruitlet (1 cm dia.), young fruitlet (12cm dia.), maturing fruit I, maturing fruit II, mature fruit), using the 'pine tree' method. Poly(A)+mRNA was isolated twice from total RNA from each stage using the Oligotex Direct mRNA kit (Qiagen). mRNA was reverse transcribed into double stranded cDNA using a modified oligo18(dT) primer with an identifying tag sequence (see table below). cDNA's from different stages were pooled in equal amounts before adaptor ligation. Tag identification when sequencing from 5' end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage 2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3 (young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4 (maturing fruit I) insert 18(A)TCGGA; Stage 5 (maturing fruit II) insert 18(A)TCGGA; Stage 6 (mature fruit) insert 18(A)TCGCT; Tag identification when sequencing from 3' end: Stage 1 (young fruitlet) CAGCA18(T) insert; Stage 2 (young fruitlet 1cm dia) CAGCA18(T) insert; Stage 3 (young fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II) TCGCA18(T) insert; Stage 6 (mature fruit) ACCGA18(T) insert. Double stranded cDNAs were size selected (more than 450 bp), adapted with EcoRI adapters at both ends and then digested with NotI. The cDNAs were then directionally cloned into EcoRI-NotI digested pBS II SK(+) phagemid vector(Stratagene). Identification of adaptors and tags in 5'-end sequenced clones:  
(Vector). . . .TAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCCTATTGTTGGG (End  
EcoRI adaptor)(Start Insert). . .AAAAAAAAAAAAAAAAAAAA(End Insert) (Start Tag)TCGGA(End Tag) (Start  
NotI site/Vector)GGGCGCCACCGCG. . . The total number of white colony forming units (cfu) in the primary library before amplification was 2.1x10<sup>6</sup> cfu (colony forming units). The background of empty clones was less than 1%. Inserts ranged from 0.5kb to 4 kb, as determined by PCR. Purified plasmid DNA from the primary library was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites flanking the cloned cDNA inserts. The purified PCR products, representing the entire cloned cDNA population, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30C. Unhybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was 5.6x10<sup>6</sup> cfu. Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAAGCTTAG 23

|||||

Db 130 AGAAGGCTGTAGAAATCTTAG 150

RESULT 40

CV657147  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CV657147 367 bp mRNA linear EST 27-OCT-2004  
Mdfrt3109h17.y1 Mdfrt Malus x domestica cDNA clone Mdfrt3109h17 5',  
mRNA sequence.  
CV657147  
CV657147.1 GI:54684533  
EST.  
Malus x domestica  
Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orvis,K.,  
Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,B., Ronko,I.,  
Tsagarisvili,R., Kennedy,S., Waterston,R. and Wilson,R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)  
Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aam06d09.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdfrt3109h17"  
/issue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdfrt"

FEATURES  
source

/note="Vector: pBluescript II SK (+); Site 1: NotI; Site 2: EcoRII; Total RNA was extracted separately from each stage (young fruitlet (<1cm), young fruitlet (1 cm dia.), young fruitlet (12cm dia.), maturing fruit I, maturing fruit II, mature fruit), using the 'pine tree' method. Poly(A)+mRNA was isolated twice from total RNA from each stage using the Oligotex Direct mRNA kit (Qiagen). mRNA was reverse transcribed into double stranded cDNA using a modified oligo18(dT) primer with an identifying tag sequence (see table below). cDNA's from different stages were pooled in equal amounts before adaptor ligation. Tag identification when sequencing from 5' end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage 2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3 (young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4 (maturing fruit I) insert 18(A)TCGGA; Stage 5 (maturing fruit II) insert 18(A)TCGGA; Stage 6 (mature fruit) insert 18(A)TCGCT; Tag identification when sequencing from 3' end: Stage 1 (young fruitlet) CAGCA18(T) insert; Stage 2 (young fruitlet 1cm dia) CAGCA18(T) insert; Stage 3 (young fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II) TCGCA18(T) insert; Stage 6 (mature fruit) ACCGA18(T) insert. Double stranded cDNAs were size selected (more than 450 bp), adapted with EcoRI adapters at both ends and then digested with NotI. The cDNAs were then directionally cloned into EcoRI-NotI digested pBS II SK(+) phagemid vector(Stratagene). Identification of adaptors and tags in 5'-end sequenced clones:  
(Vector). . . .TAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCCTATTGTTGGG (End

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 233.708 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-3  
Perfect score: 25  
Sequence: 1 caaattaagaagatttcgtgcaa 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues 9993994  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : N Geneseq\_21.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	14	ADZ75863
2	25	100.0	1731	6	ABN68276
3	25	100.0	1734	13	ADV85417
4	25	100.0	29072	13	ADV87723
5	25	100.0	29072	13	ADV78976
6	25	100.0	110000	6	ADN71527_07
7	25	100.0	110000	13	ADN81204_08
8	20.2	80.8	525	8	ACA46938
9	20.2	80.8	1731	8	ACA48261
10	18.8	75.2	1417	13	ADN45334
11	18.2	72.8	600	9	ADA31807
12	17.8	71.2	1344	8	ACA22236
13	17.8	71.2	1344	13	ADN1947
14	17.6	70.4	427	3	AAC94729
15	17.6	70.4	489	12	ADM37088
16	17.6	70.4	512	2	AAT96059
17	17.6	70.4	1173	6	ABN68462
18	17.6	70.4	1431	13	ADN18471
19	17.6	70.4	2726	13	ADO84764

20	17.6	70.4	2947	4	AAH54441	AAH54441 S. epider
21	17.6	70.4	3285	4	AAH53880	AAH53880 S. epider
22	17.6	70.4	3730	4	AAH54135	AAH54135 S. epider
23	17.6	70.4	10892	4	ABL05132	ABL05132 Drosophila
24	17.6	70.4	19922	8	ACA46526	ACA46526 Prokaryot
25	17.6	70.4	30549	6	ABN90859	ABN90859 Staphyloc
26	17.6	70.4	30612	13	ADS01031	ADS01031 Staphyloc
27	17.6	70.4	49380	4	ABL11838	ABL11838 Drosophila
28	17.6	70.4	110000	2	AAV21209_15	Continuation (16 o
29	17.6	70.4	110000	13	ABD32968_2	Continuation (3 of
30	17.6	70.4	220224	11	ACN44702	ACN44702 Human gen
31	17.6	70.4	277616	13	ABD32602	ABD32602 Human gen
32	17.6	69.6	24888	9	ADA02824	ADA02824 Human NFK
33	17.4	69.6	24888	10	ADB72362	ADB72362 Human NFK
34	17.4	69.6	24888	10	ADE95872	ADE95872 Human NFK
35	17.4	69.6	50000	10	ADC56843	ADC56843 Human IKB
36	17.4	69.6	53106	9	ADA03044	ADA03044 Human BAT
37	17.4	69.6	53106	9	ADA66328	ADA66328 Human BAT
38	17.4	69.6	53106	10	ADB72782	ADB72782 Human BAT
39	17.4	69.6	53106	11	ADL27122	ADL27122 Human gen
40	17.4	69.6	81800	6	ABK84756	ABK84756 Human cdn
41	17.4	69.6	349980	5	AAH41224	AAH41224 Pyrococcus
42	17.2	68.8	490	3	AAZ28412	AAZ28412 Human sec
43	17.2	68.8	584	8	ABZ51224	ABZ51224 Aspergill
44	17.2	68.8	2876	10	ADF37976	ADF37976 Synchroni
45	17.2	68.8	9398	14	AEA61765	AEA61765 Streptoco
46	17.2	68.8	12264	4	ABA88966	ABA88966 Escherich
47	17.2	68.8	23654	6	ABS78844	ABS78844 E. coli C
48	17.2	68.8	23654	10	ADH80411	ADH80411 Escherich
49	17.2	68.8	32132	4	AAL35943	AAL35943 Human mus
50	17.2	68.8	32132	8	ABX58931	ABX58931 cdna enco
51	17.2	68.8	32132	12	ADJ29681	ADJ29681 Human mus
52	17.2	68.8	32132	11	ACN43998_6	Continuation (7 of
53	17.2	68.8	53122	8	ABZ74619	ABZ74619 Secreted
54	17.2	68.8	58181	10	ADC21010	ADC21010 Human sec
55	17.2	68.8	58181	10	ABZ68140	ABZ68140 Human sec
56	17.2	68.8	101636	14	ADZ12890	ADZ12890 Murine ca
57	17.2	68.8	110000	2	AAV21209_10	Continuation (11 o
58	17.2	68.8	110000	11	ACN43998_5	Continuation (6 of
59	17.2	68.8	133632	11	ACN45054	ACN45054 Human gen
60	17.2	68.8	133632	11	ACN45054	ACN45054 Human gen

ALIGNMENTS

RESULT 1  
ADZ75863  
ID ADZ75863 standard; DNA; 25 BP.

XX AC ADZ75863;

XX DT 28-JUL-2005 (first entry)

XX DE Group B streptococcus phosphotransferase (pts) probe, SEQ ID NO:3.

XX KW Microorganism detection; fluorescence; diagnosis;  
streptococcus infection; infection; gynecology and obstetrics;  
phosphotransferase; probe; ss.

XX OS Streptococcus sp. 'group B'.

XX PH Key Location/Qualifiers  
FT modified\_base 25  
FT /mod\_base a  
FT /note= "3", labeled with fluorescent donor fluorescein"

XX PN US2005106578-A1.

XX PD 19-MAY-2005.

XX PF 18-NOV-2003; 2003US-00716005.

XX 18-NOV-2003; 2003US-00716005.  
PR (UHLJ/) UHL J R.  
XX (COCK/) COCKERILL F R.  
PA (AICH/) AICHINGER C.  
PA (REIS/) REISER A.  
XX Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
PI WPI; 2005-371550/38.  
XX  
DR  
XX  
XX  
PT Detecting group B streptococcus, comprises amplifying a sample with pts  
PT primers, hybridizing the sample with fluorescently labeled pts probes,  
PT and detecting the presence of fluorescence resonance energy transfer.  
XX  
PS Claim 3; SEQ ID NO 3; 13pp; English.  
XX  
XX The invention relates to a real-time PCR-based method of detecting the  
CC presence or absence of group B streptococcus (GBS) bacterial pathogens in  
CC a biological sample from an individual. The method comprises  
CC amplification of a conserved region of the phosphotransferase (pts) gene  
CC of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the  
CC amplification product with a pair of fluorescently labeled  
CC phosphotransferase probes (especially ADZ75863-ADZ75864), and detection  
CC of the presence or absence of fluorescence resonance energy transfer  
CC (FRET), where presence of FRET is indicative of the presence of GBS in  
CC the sample. GBS infection is a leading cause of neonatal morbidity and  
CC mortality, with infection occurring during childbirth. Currently, it is  
CC recommended that women are screened for GBS during week 35-37 of  
CC gestation by a culture-based method which may take up to 72 hours for a  
CC result. However, many women first present at healthcare facilities at the  
CC time of labor, and in addition, GBS infection can be transient, so that a  
CC woman free of GBS at the time of screening may not be free of GBS when  
CC she is due to give birth. The method of the invention provides a real-  
CC time assay for the detection of group B streptococcus in a sample, is  
CC more sensitive and specific than prior art non-culture based methods, and  
CC can thus be implemented for routine diagnosis of the presence of group B  
CC streptococcus. Sequences ADZ75863-ADZ75864 represent fluorescently  
CC labeled GBS phosphotransferase probes specifically claimed for use in the  
CC method of the invention.  
XX  
SQ Sequence 25 BP; 11 A; 4 C; 4 G; 6 T; 0 U; 0 Other;  
Query Match 100.0%; Score 25; DB 14; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.13; Length 25;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGCGAA 25  
DB 1 CAAATTAAGAGACTATTCTGCGAA 25  
RESULT 2  
ABN68276  
ID ABN68276 standard; DNA; 1731 BP.  
XX  
XX AC ABN68276;  
XX  
XX  
DT 01-JUL-2002 (first entry)  
XX  
XX Streptococcus polynucleotide SEQ ID NO 4465.  
DE  
XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
XX Streptococcus agalactiae.  
OS  
XX  
XX WO200234771-A2.  
PN  
XX  
PD 02-MAY-2002.  
XX  
XX

PF 29-OCT-2001; 2001WO-GB004789.  
XX  
XX 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27645.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX Claim 7; Page 3607; 4525pp; English.  
PS  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
Query Match 100.0%; Score 25; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 0.2; Length 1731;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGCGAA 25  
DB 265 CAAATTAAGAGACTATTCTGCGAA 289  
RESULT 3  
ADV85417  
ID ADV85417 standard; DNA; 1734 BP.  
XX  
XX AC ADV85417;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 6558.  
DE  
XX Antibacterial; vaccine; bacterial infection; ds.  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO200292818-A2.  
PN  
XX 21-NOV-2002.  
PD  
XX 26-APR-2002; 2002WO-IB003059.  
PF  
XX 26-APR-2001; 2001FR-00005642.  
PR  
XX (INSP ) INST PASTEUR.  
PA



PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 DR  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 4; SEQ ID NO 6558; 439pp; French.  
 XX  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX  
 XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 25; DB 13; Length 1734;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
 Db 265 CAAATTAAAGAGACTATTCTGCGAA 289  
 RESULT 4  
 ADV87723  
 ID ADV87723 standard; DNA; 29072 BP.  
 AC  
 XX ADV87723;  
 XX  
 XX 24-FEB-2005 (first entry)  
 DT  
 XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 DE  
 XX Antibacterial; Vaccine; bacterial infection; ds.  
 KW  
 XX Streptococcus agalactiae.  
 OS  
 XX FR2824074-A1.  
 PN  
 XX 31-OCT-2002.  
 PD  
 XX 26-APR-2001; 2001FR-00005642.  
 PF  
 XX 26-APR-2001; 2001FR-00005642.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA  
 XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 DR  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 1; SEQ ID NO 117; 2687pp; French.  
 XX  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transposons, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX  
 XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 25; DB 13; Length 29072;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
 Db 12160 CAAATTAAAGAGACTATTCTGCGAA 12184  
 RESULT 5  
 ADV78976  
 ID ADV78976 standard; DNA; 29072 BP.  
 AC  
 XX ADV78976;  
 XX  
 XX 24-FEB-2005 (first entry)  
 DT  
 XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 DE  
 XX Antibacterial; vaccine; bacterial infection; ds.  
 KW  
 XX Streptococcus agalactiae.  
 OS  
 XX WO200292818-A2.  
 PN  
 XX 21-NOV-2002.  
 PD  
 XX 26-APR-2002; 2002WO-IB003059.  
 PF  
 XX 26-APR-2001; 2001FR-00005642.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA  
 XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 DR  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 1; SEQ ID NO 117; 439pp; French.  
 PS  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX  
 XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
 SQ

CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (i) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (i) or (iii) are  
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
Query Match 100.0%; Score 25; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 12160 CAAATTAAAGAGACTATTCTGCGAA 12184

## RESULT 6

ABN71527\_07  
Continuation (8 of 22) of ABN71527 from base 700001 (*Streptococcus* polynucleotide SEQ ID  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession Abn71527

WP	Fragment Name	Begin	End
WP	ABN71527_00	1	110000
WP	ABN71527_01	100001	210000
WP	ABN71527_02	200001	310000
WP	ABN71527_03	300001	410000
WP	ABN71527_04	400001	510000
WP	ABN71527_05	500001	610000
WP	ABN71527_06	600001	710000
WP	ABN71527_07	700001	810000
WP	ABN71527_08	800001	910000
WP	ABN71527_09	900001	1010000
WP	ABN71527_10	1000001	1110000
WP	ABN71527_11	1100001	1210000
WP	ABN71527_12	1200001	1310000
WP	ABN71527_13	1300001	1410000
WP	ABN71527_14	1400001	1510000
WP	ABN71527_15	1500001	1610000
WP	ABN71527_16	1600001	1710000
WP	ABN71527_17	1700001	1810000
WP	ABN71527_18	1800001	1910000
WP	ABN71527_19	1900001	2010000
WP	ABN71527_20	2000001	2110000
WP	ABN71527_21	2100001	2155561

Query Match 100.0%; Score 25; DB 6; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 91030 CAAATTAAAGAGACTATTCTGCGAA 91054

## RESULT 7

ADV81204\_08  
Continuation (9 of 23) of ADV81204 from base 800001 (*Streptococcus agalactiae* complete g  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204

WP	Fragment Name	Begin	End
WP	ADV81204_00	1	110000
WP	ADV81204_01	100001	210000
WP	ADV81204_02	200001	310000
WP	ADV81204_03	300001	410000
WP	ADV81204_04	400001	510000

WP	ADV81204_05	500001	610000
WP	ADV81204_06	600001	710000
WP	ADV81204_07	700001	810000
WP	ADV81204_08	800001	910000
WP	ADV81204_09	900001	1010000
WP	ADV81204_10	1000001	1110000
WP	ADV81204_11	1100001	1210000
WP	ADV81204_12	1200001	1310000
WP	ADV81204_13	1300001	1410000
WP	ADV81204_14	1400001	1510000
WP	ADV81204_15	1500001	1610000
WP	ADV81204_16	1600001	1710000
WP	ADV81204_17	1700001	1810000
WP	ADV81204_18	1800001	1910000
WP	ADV81204_19	1900001	2010000
WP	ADV81204_20	2000001	2110000
WP	ADV81204_21	2100001	2210000
WP	ADV81204_22	2200001	2217924

Query Match 100.0%; Score 25; DB 13; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 71365 CAAATTAAAGAGACTATTCTGCGAA 71389

## RESULT 8

ACA46938  
ACA46938 standard; DNA; 525 BP.

XX	AC	ACA46938;
XX	DT	19-JUN-2003 (first entry)
XX	DE	Prokaryotic essential gene #28595.
XX	KW	Antisense; ds; prokaryotic essential gene; cell proliferation; drug design; gene.
XX	OS	<i>Staphylococcus epidermidis</i> .
XX	PN	WO200277183-A2.
XX	PD	03-OCT-2002.
XX	PF	21-MAR-2002; 2002WO-US009107.
XX	PR	21-MAR-2001; 2001US-00815242.
XX	PR	06-SEP-2001; 2001US-00948993.
XX	PR	25-OCT-2001; 2001US-0342923P.
XX	PR	08-FEB-2002; 2002US-00072851.
XX	PR	06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI: 2003-029926/02.

DR P-PSDB; ABU43068.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 34808; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 525 BP; 214 A; 71 C; 106 G; 134 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 8; Length 525;  
Best Local Similarity 88.0%; Pred. No. 27;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25  
|||||  
DB 28 CAAATTAAAGAGAAATTCGTGCAA 52

RESULT 9  
ACA48261  
ID ACA48261 standard; DNA; 1731 BP.

XX ACA48261;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #29918.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.

XX Streptococcus mutans.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU44391.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT cell for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 36131; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid;  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1731 BP; 497 A; 320 C; 399 G; 515 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 8; Length 1731;  
Best Local Similarity 88.0%; Pred. No. 30;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25

DB 265 CAGATTAAAGAGACAATTCGTACAA 289

RESULT 10

ADX45334

ID ADX45334 standard; cDNA; 1417 BP.

XX ADX45334;

XX 21-APR-2005 (first entry)

XX Plant full length insert polynucleotide seqid 20074.

XX plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.

XX Unidentified.

OS US2004034888-A1.

XX PN

```
PD 19-FEB-2004.
XX
XX
PF 28-APR-2003; 2003US-00425114.
XX
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
XX
PA (LIU//) LIU J.
PA (ZHOU//) ZHOU Y.
PA (KOVA//) KOVALIC D K.
PA (SCRE//) SCREEN S E.
PA (TABA//) TABASKA J E.
PA (CAO//) CAO Y.
XX
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
XX
XX WPI; 2004-180133/17.
XX
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX
XX Claim 1; SEQ ID NO 20074; 15pp; English.
XX
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX
XX Sequence 1417 BP; 485 A; 200 C; 247 G; 485 T; 0 U; 0 Other;
SQ
Query Match 75.2%; Score 18.8; DB 13; Length 1417;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 ATTAAGAGACTATTCTGTGCA 25
Db 926 ATTAAGAACTATTCTGTGTA 947
RESULT 11
ADA31807
ID ADA31807 standard; DNA; 600 BP.
XX
XX
XX ADA31807;
XX
XX
XX 20-NOV-2003 (first entry)
DT
XX
XX
XX DNA encoding Acinetobacter baumannii protein #3094.
DE
XX
XX
XX ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.
XX
XX
XX Acinetobacter baumannii.
OS
XX
XX
XX US6562958-B1.
PN
XX
XX
XX 13-MAY-2003.
PD
XX
XX
XX 19-FEB-2004.
XX
XX
PF 28-APR-2003; 2003US-00425114.
XX
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
XX
PA (LIU//) LIU J.
PA (ZHOU//) ZHOU Y.
PA (KOVA//) KOVALIC D K.
PA (SCRE//) SCREEN S E.
PA (TABA//) TABASKA J E.
PA (CAO//) CAO Y.
XX
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
XX
XX WPI; 2004-180133/17.
XX
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX
XX Claim 1; SEQ ID NO 20074; 15pp; English.
XX
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX
XX Sequence 1417 BP; 485 A; 200 C; 247 G; 485 T; 0 U; 0 Other;
SQ
Query Match 75.2%; Score 18.8; DB 13; Length 1417;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 ATTAAGAGACTATTCTGTGCA 25
Db 926 ATTAAGAACTATTCTGTGTA 947
RESULT 11
ADA31807
ID ADA31807 standard; DNA; 600 BP.
XX
XX
XX ADA31807;
XX
XX
XX 20-NOV-2003 (first entry)
DT
XX
XX
XX DNA encoding Acinetobacter baumannii protein #3094.
DE
XX
XX
XX ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.
XX
XX
XX Acinetobacter baumannii.
OS
XX
XX
XX US6562958-B1.
PN
XX
XX
XX 13-MAY-2003.
PD
XX
XX
XX 04-JUN-1999; 99US-00328352.
XX
XX
XX 09-JUN-1998; 98US-0088701P.
XX
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX
XX Breton G, Bush D;
XX
XX
XX WPI; 2003-576092/54.
XX
XX
XX P-PSDB; ADA35933.
XX
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
XX
XX Example; SEQ ID NO 3094; 328pp; English.
XX
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
XX
XX Sequence 600 BP; 178 A; 104 C; 120 G; 198 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 18.2; DB 9; Length 600;
Best Local Similarity 87.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCTGTGCA 24
Db 83 AAATTGAAGAACTATTCTGTA 105
RESULT 12
ACA22236
ID ACA22236 standard; DNA; 1344 BP.
XX
XX
XX ACA22236;
XX
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX
XX Prokaryotic essential gene #3893.
DE
XX
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
XX
XX Bacillus anthracis.
OS
XX
XX
XX WO200277183-A2.
PN
XX
XX
XX 03-OCT-2002.
PD
XX
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX
XX 21-MAR-2001; 2001US-00815242.
PR
XX
XX
XX 06-SEP-2001; 2001US-00948993.
PR
XX
XX
XX 25-OCT-2001; 2001US-0342923P.
PR
XX
XX
XX 08-FEB-2002; 2002US-00072851.
PR
XX
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX
XX Wang L, Zamudio C, Malone C, Haseibeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX
XX WPI; 2003-029926/02.
XX
```

DR P-PSDB; ABU18366.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX  
XX Claim 14; SEQ ID NO 10106; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1344 BP; 451 A; 213 C; 302 G; 375 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 8; Length 1344;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 AAATTAAGAGACTATTTCGTG 22  
Db 1166 AAATTAAGAGAAATTTTCGTG 1186  
|||||  
  
RESULT 13  
ADT41947  
ID ADT41947 standard; cDNA; 1344 BP.  
XX  
XX ADT41947;  
AC  
XX  
XX 02-DEC-2004 (first entry)  
DT  
XX  
XX Bacterial polynucleotide #16698.  
DE  
XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.  
XX  
XX Bacteria.  
XX  
XX US2002333675-A1.  
PN  
XX

PD 18-DEC-2003.  
XX  
XX 20-FEB-2003; 2003US-00369493.  
XX  
XX 21-FEB-2002; 2002US-0360039P.  
XX  
XX (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
PI WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 40385; 122pp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plants with  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at seqdata.uspto.gov/sequence.html.  
XX  
XX Sequence 1344 BP; 451 A; 213 C; 303 G; 377 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 13; Length 1344;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 AAATTAAGAGACTATTTCGTG 22  
Db 1166 AAATTAAGAGAAATTTTCGTG 1186  
|||||  
  
RESULT 14  
AAC94729/c  
ID AAC94729 standard; cDNA; 427 BP.  
XX  
XX AAC94729;  
AC  
XX  
XX 19-FEB-2001 (first entry)  
DT  
XX Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:1224.  
XX  
XX Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;  
KW flea infestation; vaccine; antiparasitic; therapeutic target; diagnosis;  
KW detection; ss.  
XX  
XX Ctenocephalides felis.  
OS

```

XX  WO200061621-A2.
XX
XX  19-OCT-2000.
XX
XX  07-APR-2000; 2000WO-US009437.
XX
XX  09-APR-1999; 99US-0128704P.
XX  (HESK-) HESKA CORP.
XX
XX  Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX  WPI; 2000-656323/63.
XX
XX  Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX  acids useful for the prevention, diagnosis and treatment of flea
XX  infestations.
XX
XX  Claim 26; Page 640; 964pp; English.
XX
XX  The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX  acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX  or head and nerve cord (HNC) tissue. The invention also relates to the
XX  encoded proteins. The invention additionally encompasses expression
XX  constructs, recombinant viruses and recombinant cells comprising the
XX  nucleic acids of the invention, recombinant production of the proteins,
XX  antibodies against the proteins, a method of identifying inhibitors of
XX  the proteins, and compositions comprising the inhibitors for
XX  administration to an animal. The nucleic acids, and the proteins they
XX  encode may be used in the prevention, treatment and diagnosis of diseases
XX  associated with flea infestations. For example, the nucleic acids may be
XX  used to produce an HMT or HNC protein according to standard recombinant
XX  DNA methodology by inserting the nucleic acids into a host cell and
XX  culturing the cell to express the protein. The HMT and HNC nucleic acids
XX  may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX  and quantitate the presence of cat flea or other homologous nucleic acid
XX  sequences in samples. They may also be used to study the expression and
XX  function of the proteins and their role in metabolism. The HMT and HNC
XX  proteins may be used as antigens in the production of specific
XX  antibodies, and in assays to identify modulators (agonists and
XX  antagonists) of HMT and/or HNC protein expression and activity. The anti-
XX  HMT/HNC protein antibodies and antagonists may also be used to
XX  downregulate protein expression and activity. The antibodies may also be
XX  used as diagnostic agents for detecting the presence of flea polypeptides
XX  in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX  present sequence represents a cat flea HMT cDNA of the invention
XX
XX  Sequence 427 BP; 180 A; 57 C; 38 G; 152 T; 0 U; 0 Other;
XX
XX  Query Match 70.4%; Score 17.6; DB 3; Length 427;
XX  Best Local Similarity 83.3%; Pred. No. 4.1e+02;
XX  Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY  2 AAATTAAGAGACTATTCGTGCA 25
XX  ||||| ||||| ||||| ||||| |||||
Db  359 AAATTTAAGAAACTATTCATGTA 336
XX
RESULT 15
ADM37088
ID  ADM37088 standard; cDNA; 489 BP.
XX
XX  ADM37088;
XX
XX  03-JUN-2004 (first entry)
XX
XX  Caenorhabditis elegans STARS encoding cDNA SEQ ID NO:7.
XX
XX  striated muscle activator of Rho signalin; STARS;
XX  muscle-specific actin-binding protein; cardiant; gene therapy;
XX  cardiovascular disease; cardiac hypertrophy; dilated cardiomyopathy;
XX  myocardial infarct; heart failure; drug screening; gene; ss.

```

```

XX  Caenorhabditis elegans.
XX
XX  Key Location/Qualifiers
XX  CDS 1..489
XX  /*tag= a
XX
XX  WO2004018629-A2.
XX
XX  04-MAR-2004.
XX
XX  20-AUG-2003; 2003WO-US026191.
XX
XX  20-AUG-2002; 2002US-0404706P.
XX  (TEXA ) UNIV TEXAS SYSTEM.
XX
XX  Olson E, Arai A;
XX
XX  WPI; 2004-226822/21.
XX  P-PSDB; ADM37089.
XX
XX  New striated muscle activator of Rho signalin polypeptides and nucleic
XX  acid molecules useful for preventing and treating cardiovascular
XX  diseases, e.g. cardiac hypertrophy, myocardial infarct or heart failure,
XX  or in drug screening.
XX
XX  Claim 3; SEQ ID NO 7; 127pp; English.
XX
XX  The present sequence encodes a Caenorhabditis elegans striated muscle
XX  activator of Rho signalin (STARS) protein, which is a muscle-specific
XX  actin-binding protein. Also described: (1) a polypeptide comprising any
XX  of the 5 STARS sequences of 162-399 amino acids (see SEQ ID NO:2, 4, 6, 8
XX  or 10), or a peptide comprising 10 contiguous amino acids of any of the
XX  above amino acid sequences; (2) an expression construct comprising a
XX  polynucleotide encoding a STARS polypeptide operably linked to a
XX  regulatory sequence; (3) methods of screening for modulators of STARS
XX  expression or STARS actin-binding activity; (4) a method of screening for
XX  an inhibitor of STARS-induced transcription; (5) a method of producing a
XX  STARS polypeptide in a cell; (6) a non-human transgenic animal comprising
XX  a selectable or screenable marker protein under the control of a STARS
XX  promoter, or comprising a STARS encoding nucleic acid under the control
XX  of an inducible or a constitutive promoter; (7) a method of inhibiting
XX  STARS activity; (8) methods of treating or inhibiting progression of
XX  cardiac hypertrophy, dilated cardiomyopathy, myocardial infarct or heart
XX  failure; (9) a method of preventing cardiac hypertrophy and dilated
XX  cardiomyopathy; (10) methods of increasing exercise tolerance, reducing
XX  hospitalisation, improving the quality of life or decreasing mortality
XX  and morbidity in a subject with heart failure or cardiac hypertrophy;
XX  (11) methods of producing a modulator of STARS expression or STARS actin
XX  binding activity; (12) a modulator of STARS expression identified by the
XX  method in (3); (13) an antibody that binds immunologically to STARS, or a
XX  polyclonal antibody preparation, antibodies of which bind immunologically
XX  to STARS; and (14) a hybridoma cell that produces a monoclonal antibody
XX  that binds immunologically to STARS. STARS sequences have cardiant
XX  activities, and can be used in gene therapy. Compositions and methods of
XX  the present invention are useful for preventing and treating
XX  cardiovascular diseases, such as cardiac hypertrophy, dilated
XX  cardiomyopathy, myocardial infarct or heart failure. They may also be
XX  used in drug screening applications.
XX
XX  Sequence 489 BP; 186 A; 81 C; 103 G; 119 T; 0 U; 0 Other;
XX
XX  Query Match 70.4%; Score 17.6; DB 12; Length 489;
XX  Best Local Similarity 83.3%; Pred. No. 4.1e+02;
XX  Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY  1 CAATTAAGAGACTATTCGTGCA 24
XX  ||||| ||||| ||||| ||||| |||||
Db  439 CAGCTTAAAGAGGCTATTCGAGCA 462
XX
RESULT 16

```

```
AAT96059
ID AAT96059 standard; DNA; 512 BP.
XX
AC AAT96059;
XX
XX DT 27-MAR-1998 (first entry)
XX
DE E. coli colonisation factor antigen CPA1 DNA.
XX
XX Bacterial colonisation: colonisation factor antigen; CPA1;
KW enterotoxigenic Escherichia coli; vaccine; diagnosis; research; ds.
KW
XX Escherichia coli.
OS
XX Key Location/Qualifiers
FH 1..512
FT /*tag= a
FT
XX
XX US5698416-A.
XX
XX 16-DEC-1997.
XX
XX 02-JUN-1995; 95US-00460739.
PF
XX 02-JUN-1995; 95US-00460739.
PR
XX (USSA ) US SEC OF ARMY.
PA
XX Bell BA, Wolf MK, Cassels FJ;
PI
XX WPI; 1998-051486/05.
DR
XX P-PSDB; AAW38341.
DR
XX
XX Production of bacterial colonisation factor protein - by expression under
PT control of heat-inducible promoter.
PT
XX Example 2; Col 15-16; 11pp; English.
PS
XX
XX Production of a protein that affects bacterial colonisation, comprises
CC inoculating a broth containing tryptone and yeast extract with enteric
CC bacteria containing a DNA sequence encoding the protein under the control
CC of a temperature regulated promoter, culturing the bacteria, removing the
CC bacteria from the medium and recovering the protein. The method is used
CC especially for producing the colonisation factor antigen CPA1 of
CC enterotoxigenic E. coli, i.e. the antigen encoded by the present sequence,
CC which may be used in vaccines or for diagnostic or research purposes.
CC Growing the bacteria at low temperature until the late logarithmic phase
CC increases the yield of the protein
XX
XX Sequence 512 BP; 160 A; 99 C; 104 G; 149 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 2; Length 512;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AATTTAAAGAGACTATTTCGTGCAA 25
Db 5 AATTTAAAGAGACTATTTCGTGCAA 28
RESULT 17
ABN68462
ID ABN68462 standard; DNA; 1173 BP.
XX
XX
AC ABN68462;
XX
XX 01-JUL-2002 (first entry)
DT
XX Streptococcus polynucleotide SEQ ID NO 4837.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
KW
```

```
XX Streptococcus pyogenes.
OS
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
PR
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettelin H;
PI
XX WPI; 2002-352536/38.
XX
XX P-PSDB; ABP27831.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3646; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 1173 BP; 397 A; 198 C; 237 G; 341 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 6; Length 1173;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AATTTAAAGAGACTATTTCGTGCAA 25
Db 773 AATTTAAAGAGCTAGACGTGCAA 796
RESULT 18
ADT18471/c
ID ADT18471 standard; cDNA; 1431 BP.
XX
XX
AC ADT18471;
XX
XX 13-JAN-2005 (first entry)
DT
XX Plant cDNA, Seq ID 3797.
DE
XX Plant; ss; gene; transgenic; cold tolerance; growth rate;
KW drought tolerance; disease resistance; galactomannan production;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW lignin production; extreme osmotic condition tolerance;
KW pathogens resistance; pest resistance; yield improvement; seed oil yield;
KW seed protein yield.
```

XX Viridiplantae.  
OS US2004216190-A1.  
PN 28-OCT-2004.  
XX 18-DEC-2003; 2003US-00739930.  
XX 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.  
XX (KOVA/) KOVALIC D K.  
PA Kovalic DK;  
PI WPI; 2004-757369/74.  
DR New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX Claim 1; SEQ ID NO 3797; 14pp; English.  
PS  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5344 nucleotide sequences (CDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and tape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant CDNA  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
SQ Sequence 1431 BP; 496 A; 246 C; 283 G; 406 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 1431;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTCGTGCA 25  
Db 396 AAATTAAGAACTACTCTTGAA 373

RESULT 19  
ADO84764/C  
ID ADO84764 standard; cDNA; 2726 BP.  
XX  
AC ADO84764;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 3484.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
XX 19-FEB-2004.  
XX  
XX 28-APR-2003; 2003US-00425114.  
PF  
XX 06-MAY-1999; 99US-00304517.  
PR  
XX 05-NOV-2001; 2001US-00985678.  
PR  
XX (LIU/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAO/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
PS Claim 1; SEQ ID NO 3484; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.seqdata.uspto.gov/sequence.html?DocID=2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 2726 BP; 655 A; 586 C; 662 G; 823 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 2726;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;



```
QY 1 CAAATTAAGAGACTATTTCGTGCA 24
DB 528 CAAATCAGAGACATTTTCGTGCA 505

RESULT 20
AAH54441
ID AAH54441 standard; DNA; 2947 BP.
XX AC AAH54441;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3805.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 1441-1442; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 2947 BP; 1293 A; 537 C; 448 G; 669 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 2947;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
DB 1894 CAAATAAGAGAGATAATAGTGCA 1917

RESULT 21
AAH53880
ID AAH54135 standard; DNA; 3730 BP.
XX AC AAH54135;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:3153.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 831-832; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 3285 BP; 1462 A; 580 C; 498 G; 745 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 3285;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
DB 1447 CAAATAAGAGAGATAATAGTGCA 1470

RESULT 22
AAH54135
ID AAH54135 standard; DNA; 3730 BP.
XX AC AAH54135;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:3153.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 831-832; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 3285 BP; 1462 A; 580 C; 498 G; 745 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 3285;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3499.  
XX  
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;  
KW endocarditis; ds.  
XX  
XX Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US030782.  
XX  
XX 09-NOV-1999; 99US-0164258P.  
XX  
XX (GLAX ) GLAXO GROUP LTD.  
XX  
XX PA  
XX PI Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
XX  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
XX Claim 8; Page 1071-1072; 218pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
XX Sequence 3730 BP; 1657 A; 658 C; 560 G; 855 T; 0 U; 0 Other;  
SQ  
  
Query Match 70.4%; Score 17.6; DB 4; Length 3730;  
Best Local Similarity 83.3%; Pred. No. 5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAAGAGACTATTGTCGA 24  
Db 1892 CAAATAAAGAGAAATAAGTGTGCA 1915  
  
RESULT 23  
ABJ05132  
ID ABL05132 standard; cDNA; 10892 BP.  
XX  
AC ABL05132;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 9878.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
KW  
XX Drosophila melanogaster.  
OS  
XX

PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR P-FSDB; ABB61029.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 9878; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABJ16176-ABJ30511), expressed DNA  
CC sequences (ABJ01840-ABJ16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 10892 BP; 3489 A; 2712 C; 1869 G; 2822 T; 0 U; 0 Other;  
SQ  
  
Query Match 70.4%; Score 17.6; DB 4; Length 10892;  
Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 AAATTAAAGAGACTATTGTCGCA 25  
Db 7456 AAATAAATCAGACTATTGTCGCA 7479  
  
RESULT 24  
ACA46526  
ID ACA46526 standard; DNA; 19922 BP.  
XX  
XX ACA46526;  
AC  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX  
XX Prokaryotic essential gene #28183.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
KW  
XX Staphylococcus epidermidis.  
OS  
XX WO200277183-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU42656.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 14; SEQ ID NO 34396; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 19922 BP; 8789 A; 3297 C; 2990 G; 4846 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 8; Length 19922;  
 Best Local Similarity 83.3%; Pred. No. 5.8e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 CAAATTAAAGAGACTATTCGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 1327 CAAAATAAAGAGATAATAGTGCA 1350

RESULT 25  
 ABN90859  
 ID ABN90859 standard; DNA; 30549 BP.  
 XX  
 AC ABN90859;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:322.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy; gene; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 XX USG380370-B1.  
 XX  
 XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.  
 XX  
 XX 14-AUG-1997; 97US-0055779P.  
 PR 08-NOV-1997; 97US-0064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 XX P-PSDB; ABP38314.  
 DR  
 DR Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 PT  
 XX Disclosure; SEQ ID NO 322; 267pp; English.  
 PS  
 XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 30549 BP; 13033 A; 5054 C; 4748 G; 7714 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 30549;  
 Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 CAAATTAAAGAGACTATTCGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 12031 CAAAATAAAGAGATAATAGTGCA 12054

RESULT 26  
 ADS01031  
 ID ADS01031 standard; DNA; 30612 BP.  
 XX  
 AC ADS01031;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE *Staphylococcus epidermis* polynucleotide seqid 326.  
 XX  
 KW antibacterial; vaccine; antisense therapy; *Staphylococcus epidermidis*;  
 KW recombinant expression vector; infection; computer readable medium;  
 KW computer based system; gene; ds.  
 XX  
 OS *Staphylococcus epidermidis*.  
 XX  
 XX US2004147734-A1.  
 PN  
 XX 29-JUL-2004.  
 PD  
 XX  
 XX 01-DEC-2003; 2003US-00724972.  
 PF  
 XX 08-NOV-1997; 97US-0064964P.  
 PR 13-AUG-1998; 98US-00134001.  
 PR 29-NOV-1999; 99US-00450969.  
 XX  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 XX  
 XX Doucette-Stamm L, Bush D;  
 PI WPI; 2004-580138/56.  
 DR P-PSDB; ADS04803.  
 DR

XX New isolated polypeptide and encoding nucleic acid derived from  
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
PT treating an S. epidermidis bacterial infection.  
XX  
XX Claim 5; SEQ ID NO 326; 741pp; English.  
XX  
CC The invention describes an isolated nucleic acid comprising a nucleotide  
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
CC given in the specification. Also described are: a recombinant expression  
CC vector; a cell comprising a recombinant expression vector of (1);  
CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection, comprising a nucleic acid cited above and a carrier; treating  
CC a subject for S. epidermidis infection; a recombinant or substantially  
CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
CC sample; a computer readable medium having recorded in it the nucleotide  
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
CC system for identifying fragments of the Staphylococcus genome of  
CC commercial importance; a computer based system for identifying fragments  
CC of the Staphylococcus plasmids of commercial importance; identifying  
CC commercially important nucleic acid fragments of the Staphylococcus  
CC genome and/or plasmids; and identifying an expression modulating fragment  
CC of the Staphylococcus genome and/or plasmids. The methods and  
CC compositions of the present invention are useful for the diagnosis,  
CC prevention and/or treatment of an Staphylococcal epidermidis bacterial  
CC infection. This sequence encodes a S. epidermis protein of the invention.  
XX  
SQ Sequence 30612 BP; 13069 A; 5063 C; 4756 G; 7724 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 13; Length 30612;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
DB 12031 CAAATAAAGAGATTAATAGTGCA 12054  
|||||  
  
RESULT 27  
ABL11838  
ID ABL11838 standard; cDNA; 49380 BP.  
XX  
AC ABL11838;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29996.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX

DR P-PSDB; ABB67735.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 29996; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 49380 BP; 14230 A; 10576 C; 10238 G; 14336 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 4; Length 49380;  
Best Local Similarity 83.3%; Pred. No. 6.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
DB 7624 CAAATTAAGAGCCTGTCCTGCA 7647  
|||||  
  
RESULT 28  
AAV21209\_15  
Continuation (16 of 17) of AAV21209 from base 1500001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS AAV21209 Accession AAV21209  
WP Fragment Name Begin End  
WP AAV21209\_00 1 110000  
WP AAV21209\_01 100001 210000  
WP AAV21209\_02 200001 310000  
WP AAV21209\_03 300001 410000  
WP AAV21209\_04 400001 510000  
WP AAV21209\_05 500001 610000  
WP AAV21209\_06 600001 710000  
WP AAV21209\_07 700001 810000  
WP AAV21209\_08 800001 910000  
WP AAV21209\_09 900001 1010000  
WP AAV21209\_10 1000001 1110000  
WP AAV21209\_11 1100001 1210000  
WP AAV21209\_12 1200001 1310000  
WP AAV21209\_13 1300001 1410000  
WP AAV21209\_14 1400001 1510000  
WP AAV21209\_15 1500001 1610000  
WP AAV21209\_16 1600001 1664976  
  
Query Match 70.4%; Score 17.6; DB 2; Length 110000;  
Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 AAATTTAAAGAGACTATTCTGTCAA 25  
DB 22836 AAATTTAAAGAGAAAATTAATGCAA 22859  
|||||  
  
RESULT 29  
ABD32968\_2  
Continuation (3 of 8) of ABD32968 from base 200001 (Human cancer-associated genomic DNA  
WP Sequence split into 8 fragments LOCUS ABD32968 Accession ABD32968  
WP Fragment Name Begin End  
WP ABD32968\_0 1 110000  
WP ABD32968\_1 100001 210000  
WP ABD32968\_2 200001 310000  
WP ABD32968\_3 300001 410000  
WP ABD32968\_4 400001 510000  
WP ABD32968\_5 500001 610000



```
Query Match          70.4%; Score 17.6; DB 13; Length 277616;
Best Local Similarity 83.3%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGTCGAA 25
   |||||
Db 160286 AACTTAAGAGAAATATGTGTGCAA 160309

RESULT 32
ADA02624
ID ADA02624 standard; DNA; 24888 BP.
XX
AC ADA02624;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human NFKB1L1 carcinoma associated gene, SEQ ID NO:1142.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
WPI; 2003-587068/55.
XX
New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX
Claim 1; SEQ ID NO 1142; 245pp; English.
XX
The invention relates to recombinant carcinoma associated (CA) nucleic
XX acid sequences from mouse and human (ADA01482-ADA03094), and to
XX recombinant carcinoma associated proteins (CAP) encoded by them. The
XX invention also encompasses expression vectors and host cells comprising a
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX binds to the protein, and a biochip comprising CA nucleic acid or
XX fragments thereof. The sequences of the invention were identified using
XX oncogenic retroviruses, which insert into the genome of the host organism
XX at random. Many of these do not carry transduced host oncogenes or
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX direct consequence of the effects of proviral integration into host
XX protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX leukaemia) or a propensity to carcinoma by determination of the sequence
XX of a CA gene, or by determination of CA gene expression in particular
XX tissues. CA nucleic acids, proteins and antibodies are also useful as
XX therapeutic agents and in screening and evaluating drug candidates. The
XX present sequence represents a specifically claimed human CA nucleic acid
XX sequence of the invention. Note: The complete sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;
```

```
QY 2 AAATTAAGAGACTATTTCG 20
   |||||
Db 1064 AAATTAAGAGACTATTTCG 1082

RESULT 33
ADB72362
ID ADB72362 standard; DNA; 24888 BP.
XX
AC ADB72362;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human NFKB1L1 gene.
XX
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
XX
PR 23-OCT-2001; 2001US-00004113.
XX
PR 08-NOV-2001; 2001US-00052482.
XX
PR 30-NOV-2001; 2001US-00997722.
XX
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
WPI; 2003-239337/23.
XX
New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX cancer, neoplasm, adenocarcinoma, or sarcomas.
XX
Claim 1; SEQ ID NO 190; 2304pp; English.
XX
The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;
```

Query Match 69.6%; Score 17.4; DB 10; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 2 AAATTAAGAGACTATTTCG 20
   |||||
Db 1064 AAATTAAGAGACTATTTCG 1082

RESULT 34
ADE95872
ID ADE95872 standard; DNA; 24888 BP.
XX
AC ADE95872;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human NFKB1L1 gene genomic DNA sequence.
XX
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
```

KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; NFKB1L1.  
XX Homo sapiens.  
XX WO2003039484-A2.  
XX PD 15-MAY-2003.  
XX 08-NOV-2002; 2002WO-US036071.  
XX 08-NOV-2001; 2001US-00052482.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW, Engelhard EK;  
XX WPI; 2003-441462/41.  
XX New carcinoma associated nucleic acids and proteins, useful for screening  
PT drug candidates, or for diagnosing and treating carcinomas, e.g.  
PT lymphoma, breast cancer, prostate cancer or leukemia.  
XX Claim 1; SEQ ID NO 130; 793pp; English.  
PS  
XX This invention relates to novel recombinant nucleic acids for use in  
CC diagnosis and treatment of cancer, especially carcinomas, as well as the  
CC use of compositions in screening methods. The compositions of the  
CC invention may have cytostatic activity whilst the disclosed sequences may  
CC be useful for gene therapy. The carcinoma associated nucleic acids and  
CC proteins are useful for diagnosing and treating carcinomas, for example  
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
CC drug candidates or bioactive agents capable of binding to, or modulating  
CC the activity of, a carcinoma associated protein. The present sequence is  
CC the genomic DNA sequence of the human NFKB1L1 gene which is a carcinoma  
CC associated gene of the invention.  
XX Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;  
SQ  
Query Match 69.6%; Score 17.4; DB 10; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTTCG 20  
Db 1064 AAATTAAGAGACTATTTCG 1082  
RESULT 35  
ADC56843/C  
ID ADC56843 standard; DNA; 50000 BP.  
XX  
AC ADC56843;  
XX  
XX 18-DEC-2003 (first entry)  
DT  
XX Human IKBL gene containing SNPs used to diagnose rheumatoid arthritis.  
DE  
XX rheumatoid arthritis; single nucleotide polymorphism; SNP; human; IKBL;  
KW ds.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH replace(16433,t)  
FT /\*tag= a  
FT /standard\_name= "Single nucleotide polymorphism"  
FT variation replace(15048,g)  
FT /\*tag= b  
FT /standard\_name= "Single nucleotide polymorphism"  
XX  
XX JP2002355069-A.  
PN  
XX 10-DEC-2002.  
PD

XX 10-DEC-2001; 2001JP-00376275.  
XX  
XX 12-DEC-2000; 2000JP-00378107.  
XX (INOK/) INOKO H.  
XX WPI; 2003-472592/45.  
XX Diagnosing rheumatoid arthritis by detecting a new genetic polymorphism.  
XX Claim 1; SEQ ID NO 1; 44pp; Japanese.  
XX This invention relates to a novel method of diagnosing rheumatoid  
CC arthritis by detecting single nucleotide polymorphisms (SNPs) at  
CC positions 16433, 18366, 19048. Specifically, it refers to SNPs A16433T,  
CC T18366C and C19048G. Furthermore, the present invention describes PCR  
CC primers that bind to the 3 base positions cited above and as such can be  
CC used to amplify these regions accordingly. The method is used for  
CC diagnosing rheumatoid arthritis. This polymucleotide sequence is the  
CC human IKBL gene containing SNPs of the invention.  
XX Sequence 50000 BP; 12405 A; 11455 C; 12071 G; 14069 T; 0 U; 0 Other;  
SQ  
Query Match 69.6%; Score 17.4; DB 10; Length 50000;  
Best Local Similarity 94.7%; Pred. No. 7.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTTCG 20  
Db 25310 AAATTAAGAGACTATTTCG 25292  
RESULT 36  
ADA03044/C  
ID ADA03044 standard; DNA; 53106 BP.  
XX  
XX ADA03044;  
XX  
XX 06-NOV-2003 (first entry)  
DT  
XX Human BAT1 carcinoma associated gene, SEQ ID NO:1562.  
DE  
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW Gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2003057146-A2.  
XX  
XX 17-JUL-2003.  
PD  
XX  
XX 26-DEC-2002; 2002WO-US041414.  
PF  
XX  
XX 26-DEC-2001; 2001US-00035832.  
PR  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW;  
XX WPI; 2003-587068/55.  
XX  
XX New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
PT  
XX Claim 1; SEQ ID NO 1562; 245pp; English.  
PS  
XX The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically

CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 9; Length 53106;  
 Best Local Similarity 94.7%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCG 20  
 |||||  
 Db 47351 AAATTAAGAGACTATTTCG 47333

RESULT 17  
 ADA66328/c  
 ID ADA66328 standard; DNA; 53106 BP.  
 XX  
 AC ADA66328;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human BAT1 gene genomic DNA sequence.

XX carcinoma-associated gene; CA gene; Rorc gene; MCG15938 gene; BAT1 gene;  
 KW Iqgap1 gene; IQGAP1 gene; Zpf29 gene; hCG27579 gene; Kcnj9 gene;  
 KW Kcnj9 gene; Pp3cc gene; Pp3CC gene; MCG9110 gene; hCG27579 gene;  
 KW cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;  
 KW carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;  
 KW vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;  
 KW prostate cancer; DNA vaccine; animal model; human; ds; BAT1.

XX Homo sapiens.  
 XX  
 XX WO20003053224-A2.  
 XX  
 XX PD 03-JUL-2003.

XX 20-DEC-2002; 2002WO-US041776.  
 XX  
 XX 20-DEC-2001; 2001US-00034650.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;  
 XX  
 XX WPI; 2003-569168/53.

XX Novel recombinant carcinoma-associated nucleic acid, useful for  
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing  
 PT carcinoma.

XX Claim 1; Page 67-75; 229pp; English.

XX This invention relates to a novel recombinant carcinoma-associated (CA)  
 CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences  
 CC of mouse Rorc gene or human RORC gene, mouse MCG15938 or human gene BAT1,  
 CC mouse Iqgap1 gene or human IQGAP1 gene, mouse Zpf29 gene or human

CC hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Pp3cc gene or  
 CC human PP3CC gene, mouse MCG9110 gene or human hCG27579 gene, as given in  
 CC the specification. CA genes are genes which are preferably expressed in  
 CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.  
 CC A compound which modifies the expression of the CA genes or bind to  
 CC carcinoma-associated proteins (CAP) may have cytostatic activity and the  
 CC sequences of the invention may enable the use of gene therapy or a  
 CC development of an anticancer vaccine. Therefore the invention may be  
 CC useful for diagnosis and treatment of carcinomas, especially lymphoma  
 CC carcinoma, breast cancer and prostate cancer. The CA genes may also be  
 CC useful as DNA vaccines and for generating animal models of carcinomas.  
 CC The present sequence is that of the human BAT1 gene genomic DNA sequence  
 CC of the invention.

XX Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 9; Length 53106;  
 Best Local Similarity 94.7%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCG 20  
 |||||  
 Db 47351 AAATTAAGAGACTATTTCG 47333

RESULT 38  
 ADB72782/c  
 ID ADB72782 standard; DNA; 53106 BP.

XX ADB72782;

XX 04-DEC-2003 (first entry)

XX Human BAT1 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 610; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;



```
Query Match      69.6%; Score 17.4; DB 10; Length 53106;
Best Local Similarity 94.7%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGG 20
   |||||
DB 47351 AAATTAAGAGACTATTGG 47333

RESULT 39
ADL27122/c
ID ADL27122 standard; DNA; 53106 BP.
XX
AC ADL27122;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human genomic sequence for BAT1.
XX
KW Human; ds; gene; CA gene; carcinoma associated gene; cytostatic; cancer;
KW carcinoma; lymphoma.
XX
OS Homo sapiens.
XX
PN US2003216558-A1.
XX
PD 20-NOV-2003.
XX
PF 20-DEC-2001; 2001US-00034650.
XX
PR 22-DEC-2000; 2000US-00747377.
XX
PR 02-MAR-2001; 2001US-00798586.
XX
PR
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
XX
PI Morris DW, Engelhard EK;
XX
XX
DR WPI; 2003-902052/82.
XX
XX
PT New carcinoma associated gene, useful for preparing a composition for
PT diagnosing or treating carcinoma.
PS
PS Claim 1; SEQ ID NO 10; 342pp; English.
XX
XX
CC The invention relates to a new recombinant nucleic acid (from a CA,
CC carcinoma associated, gene) appearing as ADL27113-ADL27172. Also included
CC are a host cell comprising the recombinant nucleic acid or expression
CC vector, an expression vector comprising the recombinant nucleic acid, a
CC recombinant protein (a carcinoma associated protein) comprising the
CC sequence encoded by the nucleic acid, a method for screening drug
CC candidates, a method for screening for a bioactive agent capable of
CC binding to (or modulating the activity of) a carcinoma associated
CC protein, a method for evaluating the effect of a candidate carcinoma
CC drug, a method of diagnosing carcinoma, a method for inhibiting the
CC activity of a carcinoma associated protein, a method of treating
CC carcinomas, a method of neutralising the effect of a carcinoma associated
CC protein, a biochip comprising one or more nucleic acid segments of the
CC nucleic acid, a method of diagnosing carcinoma or propensity to carcinoma
CC and a method of determining carcinoma associated gene copy number. The
CC nucleic acid is useful for preparing a composition for diagnosing or
CC treating carcinoma especially lymphomas. The present sequence is the
CC genomic sequence from a human carcinoma associated gene.
XX
XX
SQ Sequence 53106 BP; 13296 A; 11983 C; 13268 G; 14559 T; 0 U; 0 Other;

Query Match      69.6%; Score 17.4; DB 11; Length 53106;
Best Local Similarity 94.7%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGG 20
   |||||
DB 47351 AAATTAAGAGACTATTGG 47333
```

```
RESULT 40
ABK84756/c
ID ABK84756 standard; cDNA; 81800 BP.
XX
AC ABK84756;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1327.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PR
XX
PA (GENE-) GENE LOGIC INC.
XX
XX
PI Beazer-Barclay Y, Weiseman SM, Yamaga S, Vockley J;
XX
XX
DR WPI; 2002-435328/46.
XX
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
XX
PS Claim 1; SEQ ID NO 1327; 114pp; English.
XX
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
```

CC conditions. The present sequence represents a gene differentially  
CC expressed in granulocytes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 81800 BP; 19886 A; 19955 C; 20623 G; 21336 T; 0 U; 0 Other;  
Query Match 69.6%; Score 17.4; DB 6; Length 81800;  
Best Local Similarity 94.7%; Pred. No. 8.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCG 20  
DB 80932 AAATTAAGAGACTATTTCG 80914

RESULT 41  
AAH41224  
ID AAH41224 standard; DNA; 349980 BP.  
XX AC AAH41224;  
XX DT 29-OCT-2001 (first entry)  
XX DE Pyrococcus abyssi genomic fragment #3.  
XX KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.  
XX OS Pyrococcus abyssi.  
XX FH Key Location/Qualifiers  
FT misc\_feature 1..49980  
FT /tag= a  
FT /note= "This sequence overlaps with the 3' end of  
FT AAH41223"  
FT misc\_feature 30001..349980  
FT /tag= b  
FT /note= "This sequence overlaps with the 5' end of  
FT AAH41225"

XX FR2792651-A1.  
XX 27-OCT-2000.  
XX 21-APR-1999; 99FR-00005034.  
XX 21-APR-1999; 99FR-00005034.  
XX (CNRS) CNRS CENT NAT RECH SCI.  
XX (IFRS-) IFREMER INST FR RECH EXPL MER.  
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
XX Querellou J, Weissenbach J, Saurin W, Heilig R;  
XX WPI; 2001-126236/14.  
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins  
XX useful in industry.  
XX Claim 1; Page 347-443; 1657pp; French.  
XX The present invention relates to the genomic sequence of Pyrococcus  
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a  
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal  
CC vents. The present sequence is a fragment of the genomic sequence of P.  
CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223  
CC and the 3' end of this sequence overlaps with the 5' end of AAH41225. The  
CC proteins of the present invention have various potential industrial uses,  
CC since the proteins are stable at very high temperatures, come up to 110  
CC degrees centigrade. Note: This patent is in the same patent family as  
CC WO200065062, which contains additional sequences as shown in AAB99132-  
CC AAB99143, AAH75903-AAH75920 and AAG66436

XX SQ Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 U; 0 Other;  
Query Match 69.6%; Score 17.4; DB 5; Length 349980;  
Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCG 20  
DB 72176 AAATTAAGAGACTATTTCG 72194

RESULT 42  
AAC28412  
ID AAC28412 standard; cDNA; 490 BP.  
XX AC AAC28412;  
XX DT 06-OCT-2000 (first entry)  
XX DE Human secreted protein 5' EST, SEQ ID NO: 32487.  
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.  
XX OS Homo sapiens.  
XX PN EP1033401-A2.  
XX PD 06-SEP-2000.  
XX 21-FEB-2000; 2000EP-00200610.  
XX 26-FEB-1999; 99US-0122487P.  
XX (GEST) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX Claim 1; SEQ ID NO 32487; 71pp + Sequence Listing; English.  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors

XX SQ Sequence 490 BP; 136 A; 93 C; 112 G; 149 T; 0 U; 0 Other;  
Query Match 68.8%; Score 17.2; DB 3; Length 490;  
Best Local Similarity 86.4%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTCAA 25  
DB 377 ATTAAGAGACTATTTCGTCAA 398



```
PN WO2005052145-A2.
XX
XX
XX 09-JUN-2005.
XX
XX
XX 07-OCT-2004; 2004WO-US033214.
XX
XX 07-OCT-2003; 2003US-0509175P.
XX
XX (UYFL ) UNIV FLORIDA.
XX
XX Burne RA, Yi-Ywan MC;
XX
XX WPI; 2005-417990/42.
XX
XX Novel recombinant bacterial cell comprising isolated nucleic acid
PT construct, and expressing one or more alkalizing enzyme, useful for
PT treating dental caries.
XX
XX Claim 25; SEQ ID NO 2; 135pp; English.
XX
XX The invention relates to genetically engineered bacteria that promote
CC alkalization of dental biofilms. The bacteria comprise nucleic acid
CC constructs having components necessary to express functional alkali-
CC producing enzymes, such as urease, arginine deiminase and agmatine
CC deiminase. Following introduction into the oral cavity and colonization
CC of the plaque, the bacteria produce alkalis such as ammonia from urea,
CC arginine or agmatine, thereby raising the pH and reducing the incidence
CC of caries. The bacterial strain is a dental plaque-colonizing bacterium
CC such as Streptococcus mutans, Streptococcus sanguinis, Streptococcus
CC gordonii, Streptococcus parasanguis, Streptococcus vestibularis,
CC Streptococcus oralis and Streptococcus mitis. In preferred embodiments,
CC the nucleic acid construct comprises a gene cluster encoding a urease and
CC nickel transporter such as ureIABCFGDMQO, an arginine deiminase system
CC such as arcABCDTR, or an agmatine deiminase system such as aquBAC and a
CC transcriptional regulator located upstream of the agu gene cluster, or a
CC combination of these. A vector comprising the nucleic acid construct is
CC designed to target a mtl gene. The vector is inserted into the mtl gene
CC by double-crossover recombination, and the host cell is selected by its
CC inability to grow on mannitol. A composition comprising the recombinant
CC bacteria may also comprise a carrier such as a chewing gum, toothpaste or
CC lozenge. The present sequence is that of the arginine deiminase gene
CC cluster from Streptococcus gordonii strain DLI, including an anaerobic
CC regulator, flp. The gene cluster comprises open reading frames designated
CC arcA (arginine deiminase), arcB (ornithine carbamyltransferase), arcC
CC (carbamate kinase), arcD (arginine-ornithine antitransporter), arcE
CC (dipeptidase) and arcX (regulator).
XX
XX Sequence 9398 BP; 2883 A; 1709 C; 1983 G; 2823 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 14; Length 9398;
Best Local Similarity 86.4%; Pred. No. 8.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CAAATTAAGACACTATTCTGTG 22
Db 1448 CAAACTAAAGTGGCTATTCTGTG 1469
RESULT 46
ABA88966/c
ID ABA88966 standard; DNA; 12264 BP.
XX
XX ABA88966;
AC
XX
XX 11-FEB-2002 (first entry)
DT
DE
XX Escherichia coli polynucleotide SEQ ID NO 507.
XX
XX Escherichia coli; B2/D+A+; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW Pyelonephritis; antibiotic resistance; ds.
XX
```

```
OS Escherichia coli.
XX
XX WO200166572-A2.
XX
XX 13-SEP-2001.
XX
XX 12-MAR-2001; 2001WO-EP003445.
XX
XX 10-MAR-2000; 2000FR-00003145.
XX
XX 02-FEB-2001; 2001FR-00001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
XX WPI; 2001-550253/61.
XX
XX A library of DNA fragments of Escherichia coli strains for the phylogenic
PT determination of a given strain comprises polynucleotides of nature B2/D+
PT A-.
XX
XX Example 6; Fig 6; 646pp; English.
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
CC encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more frequent
CC use of broad spectrum antibiotics
XX
XX Sequence 12264 BP; 3331 A; 2836 C; 2579 G; 3518 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 4; Length 12264;
Best Local Similarity 86.4%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAGACACTATTCTGTC 23
Db 7425 AAATTAATGATACACTATTCCTGC 7404
RESULT 47
ABS78844
ID ABS78844 standard; DNA; 23654 BP.
XX
XX ABS78844;
AC
XX
XX 17-DEC-2002 (first entry)
DT
DE
XX E. coli CFT073 genomic sequence #11.
XX
XX Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KW urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; atrophic; nephrotropic; gene; ds.
XX
XX Escherichia coli.
OS
XX WO200259320-A2.
XX
XX 01-AUG-2002.
XX
XX 19-OCT-2001; 2001WO-US046833.
XX
XX 19-OCT-2000; 2000US-0242412P.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
```

XX Blattner FR, Welch RA, Burland VD;  
PI WPI; 2002-691532/74.  
XX New DNA sequences of the pathogenic *Escherichia coli* CFT073 strain,  
PT useful for preventing or treating *E. coli* CFT073 infection in humans or  
PT livestock.  
XX  
PS Claim 1; Page 17-30; 765pp; English.  
XX The present invention relates to polynucleotide sequences from the genome  
CC of the pathogenic *Escherichia coli* strain CFT073. Almost all the  
CC sequences present in *E. coli* CFT073 are absent in the previously  
CC sequenced laboratory strain K-12. The polynucleotide sequences of the  
CC invention are useful for preventing, diagnosing or treating *E. coli*  
CC CFT073 infection in humans or livestock. The polynucleotide sequences are  
CC useful for preventing urinary tract infections and pyelonephritis.  
CC Likewise, the polypeptides encoded by the different open reading frames  
CC (ORF1-5) are useful for generating a vaccine against uropathogenic *E.*  
CC *coli* strains. ABS78834-ABS79085 represent genomic sequences from *E. coli*  
CC strain CFT073  
XX  
SQ Sequence 23654 BP; 6184 A; 5664 C; 5995 G; 5806 T; 0 U; 5 Other;  
Query Match 68.8%; Score 17.2; DB 6; Length 23654;  
Best Local Similarity 86.4%; Pred. No. 9e+02; Mismatches 0; Gaps 0;  
Matches 19; Conservative 0; Indels 3; Indels 0; Gaps 0;  
QY 2 AAATTAAAGAGACTATTCTGTC 23  
||||| ||||| ||||| ||||| |||||  
Db 4839 AAATTAAAGACTATTCTGTC 4860  
RESULT 48  
ID ADH80411 standard; DNA; 23654 BP.  
XX ADH80411;  
XX 22-APR-2004 (first entry)  
DT  
DE *Escherichia coli* CFT073 genome contig #11.  
XX ds; gene; *Escherichia coli*; CFT073; *Escherichia coli* CFT073 infection.  
XX *Escherichia coli*; CFT073.  
XX US2003165870-A1.  
XX 04-SEP-2003.  
XX 01-MAR-2002; 2002US-00085959.  
XX 01-MAR-2002; 2002US-00085959.  
XX (BLAT/) BLATTNER F R.  
PA (WELC/) WELCH R A.  
PA (BURL/) BURLAND V D.  
XX Blattner FR, Welch RA, Burland VD;  
XX WPI; 2003-863698/80.  
XX New nucleic acid of *Escherichia coli* CFT073, useful for preparing a  
PT composition for diagnosing, treating or preventing infection caused by  
PT *Escherichia coli* CFT073.  
XX Claim 1; SEQ ID NO 11; 4pp; English.  
XX The invention relates to an isolated *Escherichia coli* CFT073 nucleic acid  
CC molecule. The nucleic acid is useful for preparing a composition for  
CC diagnosing, treating or preventing infection caused by *Escherichia coli*

CC CFT073. The present sequence represents a contig of the *Escherichia coli*  
XX CFT073 genome that is not present in *Escherichia coli* K-12.  
SQ Sequence 23654 BP; 6184 A; 5664 C; 5995 G; 5806 T; 0 U; 5 Other;  
Query Match 68.8%; Score 17.2; DB 10; Length 23654;  
Best Local Similarity 86.4%; Pred. No. 9e+02; Mismatches 0; Gaps 0;  
Matches 19; Conservative 0; Indels 3; Indels 0; Gaps 0;  
QY 2 AAATTAAAGAGACTATTCTGTC 23  
||||| ||||| ||||| ||||| |||||  
Db 4839 AAATTAAAGACTATTCTGTC 4860  
RESULT 49  
ID AAL35943 standard; DNA; 32132 BP.  
XX AAL35943;  
XX 08-JAN-2002 (first entry)  
DT  
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2308.  
DE  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
XX Homo sapiens.  
XX WO200155367-A1.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001338.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0244674P.  
 PR 08-NOV-2000; 2000US-0244675P.  
 PR 08-NOV-2000; 2000US-0244676P.  
 PR 08-NOV-2000; 2000US-0244677P.  
 PR 08-NOV-2000; 2000US-0244678P.  
 PR 08-NOV-2000; 2000US-0244679P.  
 PR 08-NOV-2000; 2000US-0244680P.  
 PR 08-NOV-2000; 2000US-0244681P.  
 PR 08-NOV-2000; 2000US-0244682P.  
 PR 08-NOV-2000; 2000US-0244683P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-451937/48.

PT Isolated polypeptide for treating, preventing and/ or prognosing  
 PT disorders related to the musculoskeletal system including musculoskeletal  
 PT cancers and also for testing and detection e.g. diagnosis.

XX Example 2; SEQ ID NO 2308; 781pp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins  
 CC (AB03087-AB04109) associated with the musculoskeletal system useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. The genes are isolated from a range of human tissues  
 CC disclosed in the specification. The nucleic acids, proteins, antibodies  
 CC and (ant)agonists are useful in the diagnosis, treatment and prevention  
 CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
 CC bacterial, fungal and parasitic infections. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 4; Length 32132;

Best Local Similarity 86.4%; Pred. No. 9.2e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTTAAGAGACTATTCGTGCAA 25

Db 26595 ATTTAAGAGATAATTCATGCAA 26574

RESULT 50

AAL07060/c  
ID AAL07060 standard; DNA; 32132 BP.  
XX  
AC AAL07060;  
DT 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 9748.  
DE  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155320-A2.  
XX  
XX 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225271P.  
PR 14-AUG-2000; 2000US-0225275P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234844P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.

```
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUNA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 9748; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;
XX
XX Query Match 68.8%; Score 17.2; DB 4; Length 32132;
XX Best Local Similarity 86.4%; Pred. No. 9.2e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 4 ATTAAGAGACTATTGTCGCAA 25
Db 26595 ATTAAGAGACTATTGTCGCAA 26574
XX
RESULT 51
ABX58931/c
ID ABX58931 standard; cDNA; 32132 BP.
XX
AC ABX58931;
XX
XX 26-FEB-2003 (first entry)
XX
XX cDNA encoding novel human musculoskeletal system antigen #1275.
XX
XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;
XX re-vascularisation; thrombosis; arteriosclerosis; mineral content;
XX cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
XX post-operative tissue repair; limb regeneration; neuronal growth;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX AIDS-related complex; chondrocyte growth; bone regeneration;
XX periodontal regeneration; tissue transport; bone graft; skin aging;
XX keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
XX cell growth; organ transplant; cell differentiation; body height; weight;
XX hair colour; eye colour; skin; percentage of adipose tissue;
XX pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;
XX depression; tendency for violence; pain; reproductive capability;
XX hormone level; endocrine level; appetite; libido; memory; stress;
XX storage capability; fat content; lipid content; protein content;
XX carbohydrate content; vitamin content; cofactor content;
XX nutritional component.
XX
XX Homo sapiens.
XX
XX US2002147140-A1.
XX
XX 10-OCT-2002.
XX
XX 17-JAN-2001; 2001US-00764877.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 28-JUN-2000; 2000US-0214886P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 22-AUG-2000; 2000US-0226868P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 08-SEP-2000; 2000US-0231413P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 27-SEP-2000; 2000US-0235834P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
XX 29-SEP-2000; 2000US-0236370P.
XX 02-OCT-2000; 2000US-0236802P.
XX 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
XX 02-OCT-2000; 2000US-0237039P.
XX 13-OCT-2000; 2000US-0239935P.
XX 20-OCT-2000; 2000US-0240960P.
XX 20-OCT-2000; 2000US-0241785P.
XX 20-OCT-2000; 2000US-0241809P.
XX 01-NOV-2000; 2000US-0244617P.
XX 17-NOV-2000; 2000US-0249299P.
XX 08-DEC-2000; 2000US-0251856P.
XX 08-DEC-2000; 2000US-0251868P.
XX 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-128199/12.
XX
XX Isolated nucleic acid molecules encoding musculoskeletal system
XX associated polypeptides, useful for detecting disorders, e.g. cancer.
XX
XX Disclosure; SEQ ID NO 2308; 321pp; English.
XX
XX The invention describes an isolated nucleic acid molecule comprising a
XX sequence encoding musculoskeletal system associated polypeptides useful
XX for detecting disorders, e.g., cancer or cancer metastases, in animals or
XX humans. The nucleic acid: stimulates re-vascularisation of ischaemic
XX tissues associated with conditions such as thrombosis, arteriosclerosis,
XX and other cardiovascular conditions; treats wounds due to injuries,
XX burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
XX and limb regeneration; stimulates neuronal growth; can treat and prevent
XX
```



CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or  
CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, since FGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, haematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, eye colour,  
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, circadian rhythms,  
CC depression, tendency for violence, tolerance for pain, reproductive  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140

XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 8; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 26595 ATTAAGAGATAATTCATGCAA 26574

## RESULT 52

ADJ29681/c

ID ADJ29681 standard; DNA; 32132 BP.

XX AC ADJ29681;

XX XX 20-MAY-2004 (first entry)

XX DE Human musculoskeletal system-associated genomic DNA - SEQ ID 2308.

XX KW musculoskeletal system; cytostatic; osteopathic; cancer; osteoporosis;  
XX KW gene therapy; vaccine; human; ds.

XX OS Homo sapiens.

XX XX US2004009488-A1.

XX XX 15-JAN-2004.

XX PF 13-SEP-2002; 2002US-00242515.

XX XX 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235814P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 12-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 17-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2000US-0259678P.  
PR 17-JAN-2001; 2000US-00764877.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX PI Rosen CA, Ruben SM, Barash SC;  
XX PN  
XX PD WPI; 2004-090458/09.  
XX  
PT New nucleic acid molecule, useful for preparing a medicament for  
PT preventing, treating or ameliorating a medical condition e.g., cancer of  
PT musculoskeletal tissues or osteoporosis.  
XX  
PS Disclosure; SEQ ID NO 2308; 289pp; English.  
XX  
CC The invention relates to a novel isolated musculoskeletal system-  
CC associated nucleic acid molecule. The nucleic acid of the invention  
CC demonstrates cytostatic and osteopathic activities and may be useful for  
CC preparing a medicament for preventing, treating or ameliorating a medical  
CC condition such as cancer of the musculoskeletal tissues or osteoporosis,  
CC possibly via gene therapy or vaccine production. The current sequence is  
CC that of the human musculoskeletal system-associated genomic DNA of the  
CC invention. The current sequence is not shown within the specification per  
CC se but is available on the USPTO web-site  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20040009488>.  
CC

XX  
SQ Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;  
Query Match 68.8%; Score 17.2; DB 12; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 26595 ATTAAGAGATAATTCATGCAA 26574  
|||||  
RESULT 53  
ACN43998\_6/c  
Continuation (7 of 7) of ACN43998 from base 600001 (Human genomic sequence hCG1778483.)  
WP Sequence split into 7 fragments LOCUS ACN43998 Accession ACN43998  
WP Fragment Name Begin End  
WP ACN43998\_0 1 110000  
WP ACN43998\_1 100001 210000  
WP ACN43998\_2 200001 310000  
WP ACN43998\_3 300001 410000  
WP ACN43998\_4 400001 510000  
WP ACN43998\_5 500001 610000  
WP ACN43998\_6 600001 653122  
Query Match 68.8%; Score 17.2; DB 11; Length 53122;  
Best Local Similarity 86.4%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTTCGTG 22  
|||||  
Db 7207 CAAATTAAGAGAAAATTTGTG 7186  
|||||  
RESULT 54  
ABZ74619/c  
ID ABZ74619 standard; DNA; 58181 BP.  
XX AC ABZ74619;  
XX  
DT 12-MAY-2003 (first entry)  
XX DE Secreted protein gene 372 genomic fragment HWBCN36, SEQ ID NO:1766.  
XX KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
XX acquired immunodeficiency syndrome; AIDS;  
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
XX drug screening; chromosome identification; chromosome mapping;  
XX cytostatic; gene therapy; antinflammatory; immunomodulator; anti-HIV;  
XX antianaemic; vulnery; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200277013-A2.  
XX PD 03-OCT-2002.  
XX PF 26-MAR-2002; 2002WO-US009370.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX PD WPI; 2003-040578/03.  
XX New human secreted proteins and nucleic acids, useful for detecting or  
PT treating cancer or other hyperproliferative disorders, autoimmune  
PT disorders, inflammatory disorders, HIV disease, hepatitis or anaemia.  
XX



XX PS Disclosure; Page 2272-2286; 2423pp; English.

XX CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the

XX CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,

XX CC treating or ameliorating medical conditions e.g. by protein or gene

XX CC therapy. The genes are isolated from a range of human tissues disclosed

XX CC in the specification. The nucleic acids, proteins, antibodies and

XX CC (ant)agonists are useful in the diagnosis, treatment and prevention of:

XX CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the

XX CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,

XX CC lung or urogenital; (b) immune disorders e.g. Addison's disease,

XX CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

XX CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

XX CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as

XX CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.

XX CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,

XX CC bacterial, fungal and parasitic infections

XX SQ Sequence 58181 BP; 15503 A; 12755 C; 13504 G; 16419 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 10; Length 58181;

Best Local Similarity 86.4%; Pred. No. 9,8e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25

Db 26595 ATTAAGAGACTATTCGTGCAA 26574

RESULT 57

ADZ12890/c

XX AC ADZ12890;

XX DT 16-JUN-2005 (first entry)

XX DE Murine cancer-associated genomic DNA #37.

XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;

XX KW cyostatic; gene; ds.

XX OS Mus sp.

XX PN WO2005031001-A2.

XX PD 07-APR-2005.

XX PF 23-SEP-2004; 2004WO-US031617.

XX PR 23-SEP-2003; 2003US-00669920.

XX PA (CHIR ) CHIRON CORP.

XX PI Morris DW, Malandro MS;

XX PI WPI; 2005-273395/28.

XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,

XX PT comprises two or more nucleic acid probes.

XX PS Disclosure; SEQ ID NO 410; 198pp; English.

XX CC The invention relates to a nucleic acid array for detecting a cancer

XX CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.

XX CC The invention also relates to a peptide array comprising two or more

XX CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound

XX CC that binds to a polypeptide, an isolated antibody or its fragment which

XX CC binds to a polypeptide, which is prepared by immunizing a host animal

XX CC with a composition comprising the polypeptide or its antigen binding

XX CC fragment and collecting cells from the host expressing antibodies against

XX CC the antigen or its antigen binding fragment, a composition comprising the

CC antibody and a carrier, a method of screening for anticancer activity, a

CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a

CC method of treating cancer and a method of inhibiting expression of a CA

CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA

CC nucleic acids. The antibody is useful for detecting the presence or

CC absence of cancer cells in an individual which involves contacting cells

CC from the individual with the antibody and detecting a complex of a CA

CC protein from the cancer cells and the antibody, where the detection of

CC the complex correlates with the presence of cancer cells in the

CC individual. The composition is useful for inhibiting growth of cancer

CC cells in an individual or for delivering a therapeutic agent to cancer

CC cells in an individual. The invention is also useful for diagnosing

CC cancer, for treating cancer and for inhibiting expression of a CA gene in

CC a cell. This sequence represents murine cancer-associated genomic DNA of

XX CC the invention.

SQ Sequence 101636 BP; 28997 A; 21321 C; 20719 G; 29584 T; 0 U; 1015 Other;

Query Match 68.8%; Score 17.2; DB 14; Length 101636;

Best Local Similarity 86.4%; Pred. No. 1e+03; 3; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTG 22

Db 25 CAACTTAAGAGACTATTCGTG 4

RESULT 58

AAV21209 10

Continuation (11 of 17) of AAV21209 from base 1000001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209

WP Fragment Name Begin End

WP AAV21209\_00 1 110000

WP AAV21209\_01 100001 210000

WP AAV21209\_02 200001 310000

WP AAV21209\_03 300001 410000

WP AAV21209\_04 400001 510000

WP AAV21209\_05 500001 610000

WP AAV21209\_06 600001 710000

WP AAV21209\_07 700001 810000

WP AAV21209\_08 800001 910000

WP AAV21209\_09 900001 1010000

WP AAV21209\_10 1000001 1110000

WP AAV21209\_11 1100001 1210000

WP AAV21209\_12 1200001 1310000

WP AAV21209\_13 1300001 1410000

WP AAV21209\_14 1400001 1510000

WP AAV21209\_15 1500001 1610000

WP AAV21209\_16 1600001 1664976

Query Match 68.8%; Score 17.2; DB 2; Length 110000;

Best Local Similarity 86.4%; Pred. No. 1e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25

Db 49773 ATTAAGAGACTATTCGTGCAA 49794

RESULT 59

ACN43998\_5/c

Continuation (6 of 7) of ACN43998 from base 500001 (Human genomic sequence HCG1778483.)

WP Sequence split into 7 fragments LOCUS ACN43998 Accession Acn43998

WP Fragment Name Begin End

WP ACN43998\_0 1 110000

WP ACN43998\_1 100001 210000

WP ACN43998\_2 200001 310000

WP ACN43998\_3 300001 410000

WP ACN43998\_4 400001 510000

WP ACN43998\_5 500001 610000

WP ACN43998\_6 600001 653122

Query Match 68.8%; Score 17.2; DB 11; Length 110000;

Best Local Similarity 86.4%; Pred. No. 1e+03; Mismatches 3; Indels 0; Gaps 0;  
Matches 19; Conservative 0;

QY 1 CAAATTAAAGAGACTATTCGTG 22  
DB 107207 CAAATTAAAGAGAAATTTGTG 107186

RESULT 60

ACN45054  
ID ACN45054 standard; DNA; 133632 BP.

XX AC ACN45054;  
XX

DT 18-NOV-2004 (first entry)

XX Human genomic sequence hCG28560.

DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX

XX Homo sapiens.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
PT comprises a nucleotide sequence.

XX Claim 1; SEQ ID NO 1810; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to carcinoma Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published

XX SQ Sequence 133632 BP; 39756 A; 25635 C; 26723 G; 41518 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 11; Length 133632;  
Best Local Similarity 86.4%; Pred. No. 1e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTCGTGCA 24  
DB 22221 AATAAGGAGACTATTCATGCA 22242

Search completed: January 27, 2006, 22:15:30  
Job time : 243.708 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 186.966 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-4  
Perfect score: 20  
Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues 9993994  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : N\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001as.\*  
5: Geneseq2001bs.\*  
6: Geneseq2002as.\*  
7: Geneseq2002bs.\*  
8: Geneseq2003as.\*  
9: Geneseq2003bs.\*  
10: Geneseq2003cs.\*  
11: Geneseq2003ds.\*  
12: Geneseq2004as.\*  
13: Geneseq2004bs.\*  
14: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	14	ADZ75864
2	20	100.0	1731	6	ABN68276
3	20	100.0	1734	13	ADV85417
4	20	100.0	29072	13	ADV87723
5	20	100.0	29072	13	ADV78976
6	20	100.0	110000	6	ABN71527_07
7	20	100.0	110000	13	ADV81204_08
8	18.4	92.0	9473	1	AA92768
9	17.4	87.0	272	4	AA547567
10	17.4	87.0	272	8	ACA12260
11	17.4	87.0	732	8	ACA18397
12	17.4	87.0	4911	4	ABL10841
13	17.4	87.0	7886	4	ABL10840
14	17.4	87.0	99090	12	ADM72250
15	17.4	87.0	177587	11	ACN44806
16	17.4	87.0	177930	14	ADZ13195
17	17	85.0	1081	2	AA113710
18	17	85.0	1081	6	ABS99505
19	17	85.0	22467	4	AA533425

C	20	17	85.0	22467	4	AA533423
C	21	16.8	84.0	310	2	AAQ77281
C	22	16.8	84.0	1119	3	AAC46080
C	23	16.8	84.0	2689	2	AA505230
C	24	16.8	84.0	5626	5	ABA19882
C	25	16.8	84.0	9360	2	AAQ02830
C	26	16.8	84.0	9630	3	AAZ89828
C	27	16.8	84.0	9633	1	AA808890
C	28	16.8	84.0	9643	1	AA808859
C	29	16.8	84.0	9663	3	AA819388
C	30	16.8	84.0	9670	12	ADL23697
C	31	16.8	84.0	9672	2	AAQ20616
C	32	16.8	84.0	9693	1	AA922119
C	33	16.8	84.0	9726	3	AA819377
C	34	16.8	84.0	9822	1	AA91774
C	35	16.8	84.0	14425	10	ADC87184
C	36	16.8	84.0	16270	10	ADC87444
C	37	16.8	84.0	41079	13	ABD33401
C	38	16.8	84.0	41104	14	ADZ13504
C	39	16.8	84.0	80268	13	ABD32951
C	40	16.4	82.0	453	2	AAT83759
C	41	16.4	82.0	469	13	ADQ55786
C	42	16.4	82.0	507	10	ABX08116
C	43	16.4	82.0	507	12	ADM91806
C	44	16.4	82.0	512	12	ACH75526
C	45	16.4	82.0	626	13	ACN52635
C	46	16.4	82.0	711	8	ACF73876
C	47	16.4	82.0	714	6	ABN93331
C	48	16.4	82.0	714	13	ADS04339
C	49	16.4	82.0	762	8	ACA24380
C	50	16.4	82.0	924	2	AAV75056
C	51	16.4	82.0	1035	2	AAV43303
C	52	16.4	82.0	1110	2	AAV43316
C	53	16.4	82.0	1196	3	AAA26350
C	54	16.4	82.0	1196	12	ADL71411
C	55	16.4	82.0	1638	2	AAV68066
C	56	16.4	82.0	1808	6	ABQ54797
C	57	16.4	82.0	2014	6	ABA90791
C	58	16.4	82.0	2019	6	ABA90848
C	59	16.4	82.0	3807	6	ABQ67188
C	60	16.4	82.0	3912	8	ACA22058

ALIGNMENTS

RESULT 1  
ADZ75864

ID ADZ75864 standard; DNA; 20 BP.

XX AC ADZ75864;

XX DT 28-JUL-2005 (first entry)

XX DE Group B streptococcus phosphotransferase (pts) probe, SEQ ID NO:4.

XX KW Microorganism detection; fluorescence; diagnosis;

XX KW streptococcus infection; infection; gynecology and obstetrics;

XX KW phosphotransferase; probe; ss.

XX OS Streptococcus sp. 'group B'.

XX Key Location/Qualifiers

XX FT modified\_base 1 /\*tag= a

XX FT /mod\_base= OTHER

XX FT /note= "5' labeled with fluorescent acceptor LC-Red640"

XX US US2005106578-A1.

XX PD 19-MAY-2005.

XX PF 18-NOV-2003; 2003US-00716005.

```
XX 18-NOV-2003; 2003US-00716005.
PR (UHLJ/) UHL J R.
XX (COCK/) COCKERILL F R.
PA (AICH/) AICHINGER C.
PA (REIS/) REISER A.
XX
PI Uhl JR, Cockerill FR, Aichinger C, Reiser A;
XX WPI; 2005-371550/38.
DR
XX
XX Detecting group B streptococcus, comprises amplifying a sample with pts
PT primers, hybridizing the sample with fluorescently labeled pts probes,
PT and detecting the presence of fluorescence resonance energy transfer.
XX
XX Claim 3; SEQ ID NO 4; 13pp; English.
XX
XX The invention relates to a real-time PCR-based method of detecting the
CC presence or absence of group B streptococcus (GBS) bacterial pathogens in
CC a biological sample from an individual. The method comprises
CC amplification of a conserved region of the phosphotransferase (pts) gene
CC of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the
CC amplification product with a pair of fluorescently labeled
CC phosphotransferase probes (especially ADZ75863-ADZ75864), and detection
CC of the presence or absence of fluorescence resonance energy transfer
CC (FRET), where presence of FRET is indicative of the presence of GBS in
CC the sample. GBS infection is a leading cause of neonatal morbidity and
CC mortality, with infection occurring during childbirth. Currently, it is
CC recommended that women are screened for GBS during week 35-37 of
CC gestation by a culture-based method which may take up to 72 hours for a
CC result. However, many women first present at healthcare facilities at the
CC time of labor, and in addition, GBS infection can be transient, so that a
CC woman free of GBS at the time of screening may not be free of GBS when
CC she is due to give birth. The method of the invention provides a real-
CC time assay for the detection of group B streptococcus in a sample, is
CC more sensitive and specific than prior art non-culture based methods, and
CC can thus be implemented for routine diagnosis of the presence of group B
CC streptococcus. Sequences ADZ75863-ADZ75864 represent fluorescently
CC labeled GBS phosphotransferase probes specifically claimed for use in the
CC method of the invention.
XX
XX Sequence 20 BP; 10 A; 3 C; 5 G; 2 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 20; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAAACAGG 20
DB 1 CAAGTAAATGCAGAAACAGG 20
RESULT 2
ABN68276
ID ABN68276 standard; DNA; 1731 BP.
XX
AC ABN68276;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 4465.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
PD 02-MAY-2002.
XX
```

```
PF 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR P-FSDB; ABP27645.
DR
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3607; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 20; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAAACAGG 20
DB 292 CAAGTAAATGCAGAAACAGG 311
RESULT 3
ADV85417
ID ADV85417 standard; DNA; 1734 BP.
XX
XX ADV85417;
AC
XX
XX 24-FEB-2005 (first entry)
DT
XX
XX Streptococcus agalactiae DNA sequence, SEQ ID 6558.
DE
XX Antibacterial; vaccine; bacterial infection; ds.
KW Streptococcus agalactiae.
OS
XX WO200292818-A2.
PN
XX 21-NOV-2002.
PD
XX 26-APR-2002; 2002WO-IB003059.
PF
XX 26-APR-2001; 2001FR-00005642.
PR
XX (INSP ) INST PASTEUR.
PA
```



PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
and identification of therapeutic targets.

XX

XX Claim 1; SEQ ID NO 117; 2687pp; French.

XX

CC The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
CC agalactiae involved in the synthesis of amino acids, cell membranes, and  
CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
CC phospholipid metabolism, nucleotide metabolism including purines,  
CC pyrimidines and/or nucleosides, regulatory functions, replication,  
CC transcription, translation, protein transport, adaptation to atypical  
CC conditions, sensitivity to medicines and/or analogues, functions related  
CC to transposons, biosynthesis of cofactors, prosthetic groups and  
CC transporters, cell membrane proteins and cellular machinery. (I) are  
CC useful for the detection and/or amplification of nucleic acids.  
CC Pharmaceutical composition comprising (I) or (II) are useful for  
CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.

XX

XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

QQ

Query Match 100.0%; Score 20; DB 13; Length 29072;  
Best Local Similarity 100.0%; Prod. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 CAAGTAATGCGAGAACAGG 20  
| | | | | | | | | | | | | | | |  
DB 12187 CAAGTAATGCGAGAACAGG 12206

RESULT 5

ADV78976

ID ADV78976 standard; DNA; 29072 BP.

XX

AC ADV78976;

AC

XX

DT 24-FEB-2005 (first entry)

XX

DE Streptococcus agalactiae DNA sequence, SEQ ID 117.

XX

KW Antibacterial; vaccine; bacterial infection; ds.

OS Streptococcus agalactiae.

XX

XX WO200292818-A2.

PN

XX

PD 21-NOV-2002.

XX

PF 26-APR-2002; 2002WO-IB003059.

XX

PR 26-APR-2001; 2001FR-00005642.

XX

XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.

PA

PA

PI Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalioui L;  
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
DR WPI; 2004-101891/11.

XX

PT Genomic nucleotide sequences encoding polypeptides of Streptococcus  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
and identification of therapeutic targets.

XX

XX Claim 1; SEQ ID NO 117; 439pp; French.

XX

CC The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The

CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 12187 CAAGTAAATGCAGAAACAGG 12206

#### RESULT 6

ABN71527\_07  
Continuation (8 of 22) of ABN71527 from base 700001 (Streptococcus polynucleotide SEQ ID  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession Abn71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	1000001	1110000
WP ABN71527_11	1100001	1210000
WP ABN71527_12	1200001	1310000
WP ABN71527_13	1300001	1410000
WP ABN71527_14	1400001	1510000
WP ABN71527_15	1500001	1610000
WP ABN71527_16	1600001	1710000
WP ABN71527_17	1700001	1810000
WP ABN71527_18	1800001	1910000
WP ABN71527_19	1900001	2010000
WP ABN71527_20	2000001	2110000
WP ABN71527_21	2100001	2155561

Query Match 100.0%; Score 20; DB 6; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 91057 CAAGTAAATGCAGAAACAGG 91076

#### RESULT 7

ADV81204\_08  
Continuation (9 of 23) of ADV81204 from base 800001 (Streptococcus agalactiae complete  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204

Fragment Name	Begin	End
WP ADV81204_00	1	110000
WP ADV81204_01	100001	210000
WP ADV81204_02	200001	310000
WP ADV81204_03	300001	410000
WP ADV81204_04	400001	510000

WP	ADV81204_05	500001	610000
WP	ADV81204_06	600001	710000
WP	ADV81204_07	700001	810000
WP	ADV81204_08	800001	910000
WP	ADV81204_09	900001	1010000
WP	ADV81204_10	1000001	1110000
WP	ADV81204_11	1100001	1210000
WP	ADV81204_12	1200001	1310000
WP	ADV81204_13	1300001	1410000
WP	ADV81204_14	1400001	1510000
WP	ADV81204_15	1500001	1610000
WP	ADV81204_16	1600001	1710000
WP	ADV81204_17	1700001	1810000
WP	ADV81204_18	1800001	1910000
WP	ADV81204_19	1900001	2010000
WP	ADV81204_20	2000001	2110000
WP	ADV81204_21	2100001	2210000
WP	ADV81204_22	2200001	2217924

Query Match 100.0%; Score 20; DB 13; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 71392 CAAGTAAATGCAGAAACAGG 71411

#### RESULT 8

AAN92768

ID AAN92768 standard; DNA; 9473 BP.

XX

AC AAN92768;

XX

DT 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 18-MAY-1990 (first entry)

XX

DE HIV-2 variant HIV-D194 clone.

XX

KW HIV-2; AIDS; HIV-D194; ss.

XX

OS Human immunodeficiency virus 2.

XX

PN EP347365-A.

XX

PD 20-DEC-1989.

XX

PF 13-JUN-1989; 89EP-00710057.

XX

PR 14-JUN-1988; 88DE-03820223.

XX

PA (DIAG-) DIAGEN INST MOLEKUL.

XX

PI (CHEM-) CHEMOTHERAPEUT GEORG-SPE.

XX

PI Henco K, Vombriesen H, Immanuelmann A, Kuhnel H, Dietrich U;

PI Rubsamenwa H, Adamski M;

XX

DR WPI; 1989-372573/51.

XX

XX

PT New HIV-2 virus variant isolates - used to obtain prods. for identifying

PT infections by HIV-2 or for vaccines for protection against AIDS

XX

PS Disclosure; Fig 4; 27pp; English.

XX

CC The sequence characterises the clone HIV-D194, it starts with R/U5 region

CC of the LTR and ends with the U5 region. The DNA can be used to raise Abs

CC for serological detection, and polypeptides encoded by the sequence used

CC in vaccination. (Updated on 25-MAR-2003 to correct PF field.) (Updated on

CC 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI

CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX

SQ Sequence 9473 BP; 3215 A; 1953 C; 2374 G; 1923 T; 0 U; 8 Other;  
Query Match 92.0%; Score 18.4; DB 1; Length 9473;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAGTAAATGCAGAAACAGG 20  
|||||  
DB 4213 CAAAGTAAATGCAGAAATAGG 4232

RESULT 9  
AAS47567/c  
ID AAS47567 standard; DNA; 272 BP.  
XX  
AC AAS47567;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Enterococcus faecalis cellular proliferation inhibitory sequence #137.  
XX  
KW Antisense; ss; prokaryotic cellular proliferation; antibiotic;  
KW antibacterial; drug design.  
XX  
OS Enterococcus faecalis.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US009180.  
XX  
PR 21-MAR-2000; 2000US-0191078P.  
PR 23-MAY-2000; 2000US-0206848P.  
PR 26-MAY-2000; 2000US-0207727P.  
PR 23-OCT-2000; 2000US-0242578P.  
PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
PR 16-FEB-2001; 2001US-0269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
XX  
PS Claim 1; SEQ ID NO 144; 51lpp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence is an antisense oligonucleotide of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 272 BP; 74 A; 60 C; 40 G; 98 T; 0 U; 0 Other;

Query Match 87.0%; Score 17.4; DB 4; Length 272;

Best Local Similarity 94.7%; Pred. No. 4.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAGTAAATGCAGAAACAG 19  
|||||  
DB 166 CAAAGTAAATGCAGAAACAG 148

RESULT 10  
ACA12260/c  
ID ACA12260 standard; DNA; 272 BP.  
XX  
AC ACA12260;  
XX  
DT 27-OCT-2003 (revised)  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene antisense oligonucleotide #130.  
XX  
KW Antisense; ss; prokaryotic essential gene; cell proliferation;  
KW drug design.  
XX  
OS Archaea.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 1; SEQ ID NO 130; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the 6213  
 CC antisense sequences of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 272 BP; 74 A; 60 C; 40 G; 98 T; 0 U; 0 Other;  
 Query Match 87.0%; Score 17.4; DB 8; Length 272;  
 Best Local Similarity 94.7%; Pred. No. 4.3e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CAAGTAAATGCAGAAACAG 19  
 Db 166 CAAGTAAATGAAGAAACAG 148  
 |||||  
 |||||  
 RESULT 11  
 ACA18397  
 ID ACA18397 standard; DNA; 732 BP.  
 XX  
 AC ACA18397;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #54.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU14527.  
 XX  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 6267; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 732 BP; 250 A; 121 C; 169 G; 192 T; 0 U; 0 Other;  
 Query Match 87.0%; Score 17.4; DB 8; Length 732;  
 Best Local Similarity 94.7%; Pred. No. 4.7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CAAGTAAATGCAGAAACAG 19  
 Db 118 CAAGTAAATGAAGAAACAG 136  
 |||||  
 |||||  
 RESULT 12  
 ABL10841/c  
 ID ABL10841 standard; cDNA; 4911 BP.  
 XX  
 AC ABL10841;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27005.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB66738.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Claim 1; SEQ ID NO 27005; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4911 BP; 1375 A; 1371 C; 1197 G; 968 T; 0 U; 0 Other;

Query Match 87.0%; Score 17.4; DB 4; Length 4911;  
Best Local Similarity 94.7%; Pred. No. 5.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1405 AGGTAATGCAGAAACAGG 1387

RESULT 13  
ID ABL10840/c  
ID ABL10840 standard; cDNA; 7886 BP.  
XX  
AC ABL10840;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27002.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PI WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB66737.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 27002; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 7886 BP; 2403 A; 1939 C; 1717 G; 1827 T; 0 U; 0 Other;

Query Match 87.0%; Score 17.4; DB 4; Length 7886;  
Best Local Similarity 94.7%; Pred. No. 5.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
| | | | | | | | | | | | | | | | | | | | | |  
XX

Db 3324 AGGTAATGCAGAAACAGG 3306  
RESULT 14  
ID ADM72250  
ID ADM72250 standard; DNA; 99090 BP.  
XX  
AC ADM72250;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE O. minuta Pi2 region contiguous sequence.  
XX  
KW Pi2; NBS1; plant protectant; gene therapy; rice; disease resistance;  
KW gene; ds.  
XX  
OS Oryza minuta.  
XX  
PN WO2004022715-A2.  
XX  
PD 18-MAR-2004.  
XX  
PF 08-SEP-2003; 2003WO-US027913.  
XX  
PR 09-SEP-2002; 2002US-0409216P.  
PR 18-MAR-2003; 2003US-0455713P.  
PR 05-SEP-2003; 2003US-00656394.  
XX  
PA (OHIS ) UNIV OHIO STATE.  
XX  
PI Wang G;  
XX  
XX WPI; 2004-257576/24.  
XX  
PT New rice Pi2-like disease resistance nucleic acid molecule that confers  
PT disease resistance in plants, useful for creating or enhancing disease  
PT resistance in plants.  
XX  
PS Example 3; SEQ ID NO 13; 120pp; English.  
XX  
XX The invention relates to novel broad-spectrum resistance gene Pi2 and the  
CC NBS(1-6) genes present in the Pi2 gene cluster region. The rice Pi2-like  
CC disease resistance nucleic acid molecules are useful for enhancing  
CC disease resistance in plants. The present sequence represents a Pi2  
CC region contiguous sequence  
XX  
SQ Sequence 99090 BP; 28990 A; 20403 C; 21246 G; 28450 T; 0 U; 1 Other;

Query Match 87.0%; Score 17.4; DB 12; Length 99090;  
Best Local Similarity 94.7%; Pred. No. 7.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 32637 AAGTAAATGCAGAAACAGG 32655

RESULT 15  
ACN44806  
ID ACN44806 standard; DNA; 177587 BP.  
XX  
AC ACN44806;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human genomic sequence hCG40093.  
XX  
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2003073826-A2.  
XX

```
PD 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX
XX 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX
XX WPI; 2003-328604/31.
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 1438; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published
XX
XX SQ Sequence 177587 BP; 49045 A; 38259 C; 39386 G; 50877 T; 0 U; 20 Other;
XX
XX Query Match 87.0%; Score 17.4; DB 11; Length 177587;
XX Best Local Similarity 94.7%; Pred. No. 7.5e+02;
XX Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 CAAGTAAATGCAGAAACAG 19
DB 102050 CAAGTAAATTCAGAAACAG 102068
XX
XX RESULT 16
XX ADZ13195
XX ID ADZ13195 standard; DNA; 177930 BP.
XX
XX AC ADZ13195;
XX
XX DT 16-JUN-2005 (first entry)
XX
XX DE Human cancer-associated genomic DNA #58.
XX
XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX cytostatic; gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO2005031001-A2.
XX
XX PD 07-APR-2005.
XX
XX PF 23-SEP-2004; 2004WO-US031617.
XX
XX PR 23-SEP-2003; 2003US-00669920.
XX
XX PA (CHIR ) CHIRON CORP.
XX
XX PI Morris DW, Malandro MS;
XX
XX DR WPI; 2005-273395/28.
XX
XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,
XX comprises two or more nucleic acid probes.
XX
XX PS Disclosure; SEQ ID NO 715; 198pp; English.
XX
XX The invention relates to a nucleic acid array for detecting a cancer
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.
XX The invention also relates to a peptide array comprising two or more
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound
XX that binds to a polypeptide, an isolated antibody or its fragment which
XX binds to a polypeptide, which is prepared by immunizing a host animal
XX with a composition comprising the polypeptide or its antigen binding
XX fragment and collecting cells from the host expressing antibodies against
XX the antigen or its antigen binding fragment, a composition comprising the
XX antibody and a carrier, a method of screening for anticancer activity, a
XX method of detecting a CA nucleic acid, a method of diagnosing cancer, a
XX method of treating cancer and a method of inhibiting expression of a CA
XX nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
XX absence of cancer cells in an individual which involves contacting cells
XX from the individual with the antibody and detecting a complex of a CA
XX protein from the cancer cells and the antibody, where the detection of
XX the complex correlates with the presence of cancer cells in the
XX individual. The composition is useful for inhibiting growth of cancer
XX cells in an individual or for delivering a therapeutic agent to cancer
XX cells in an individual. The invention is also useful for diagnosing
XX cancer, for treating cancer and for inhibiting expression of a CA gene in
XX a cell. This sequence represents human cancer-associated genomic DNA of
XX the invention.
XX
XX SQ Sequence 177930 BP; 49132 A; 38321 C; 39437 G; 51020 T; 0 U; 20 Other;
XX
XX Query Match 87.0%; Score 17.4; DB 14; Length 177930;
XX Best Local Similarity 94.7%; Pred. No. 7.5e+02;
XX Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 CAAGTAAATGCAGAAACAG 19
DB 102050 CAAGTAAATTCAGAAACAG 102068
XX
XX RESULT 17
XX AAX13710/c
XX ID AAX13710 standard; DNA; 1081 BP.
XX
XX AC AAX13710;
XX
XX DT 19-MAR-1999 (first entry)
XX
XX DE Enterococcus faecalis genome contig SEQ ID NO:773.
XX
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
XX OS Enterococcus faecalis.
XX
XX PN WO9850555-A2.
XX
XX PD 12-NOV-1998.
XX
XX PF 04-MAY-1998; 98WO-US008985.
XX
XX PR 06-MAY-1997; 97US-0044031P.
XX
XX PR 16-MAY-1997; 97US-004655P.
XX
XX PR 14-NOV-1997; 97US-0066009P.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Kunsch CA, Dillon PJ, Barash SC;
XX
XX DR WPI; 1999-045171/04.
XX
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides -
XX used to develop products for the detection of Enterococcus and for use in
```

PT vaccines for prevention or attenuation of Enterococcus infection.

XX Claim 1; Page 1979-1980; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC AAX12938 to AAX13919 represent these nucleotide sequences which are

CC primary nucleotide sequences, also known as contigs. The computer-based

CC system can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence of

CC Enterococcus faecalis in samples. They can also be used for diagnosing

CC Enterococcal infection in an animal and monitoring progression of

CC disease, and for identifying agents which can be used to modulate the

CC growth or pathogenicity of Enterococcus faecalis, or another related

CC organism, in vivo or in vitro. In particular the polypeptides encoded by

CC the Enterococcus faecalis nucleotide sequences can be used in vaccines to

XX prevent or attenuate an Enterococcal infection

SQ Sequence 1081 BP; 310 A; 237 C; 183 G; 334 T; 0 U; 17 Other;

Query Match 85.0%; Score 17; DB 2; Length 1081;

Best Local Similarity 89.5%; Pred. No. 7.3e+02;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CAAGTAAATGCAGAAACAG 19

Db 608 CAAGTAAATGCAGAAACAG 590

||||||| :|||||

RESULT 18

ABS99505/c ,

ID ABS99505 standard; DNA; 1081 BP.

XX ABS99505;

AC ABS99505;

XX 18-DEC-2002 (first entry)

DT Enterococcus faecalis contig sequence #773.

DE

XX Computer readable medium; Enterococcus faecalis; microbe; growth;

KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;

KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;

KW biochip technology; antibacterial; modulator of nucleic acid expression;

KW contig; ds.

XX Enterococcus faecalis.

OS US2002120116-A1.

XX 29-AUG-2002.

PD 04-MAY-1998; 98US-00070927.

PF 04-MAY-1998; 98US-00070927.

XX (KUNS/) KUNSCH C A.

PA (DILL/) DILLON P J.

PA (BARA/) BARASH S.

XX Kunsch CA, Dillon PJ, Barash S;

PI WPI; 2002-750065/81.

XX

XX Computer readable medium having recorded on it a Enterococcus faecalis

PT nucleotide sequence useful for detecting diseases related to Enterococcus

PT infections in animals.

XX Claim 1; Page; 119pp; English.

XX The present invention relates to a new computer readable medium with an

CC Enterococcus faecalis nucleotide sequence. The invention is useful to

CC diagnose the presence of E.faecalis in a sample or determining the

CC presence of a specific microbe in a sample. The invention is also useful

CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine to

CC confer resistance to Enterococcal infection, for commercial, therapeutic

CC and industrial purposes, and for fermenting a particular sugar source or

CC to produce a particular metabolite. The invention is useful for detecting

CC diseases related to Enterococcus infections in animals, and for detecting

CC E.faecalis using biochip technology. The present nucleic acid sequence

CC represents an Enterococcus faecalis contig DNA sequence of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format directly from USPTO

CC at http.sequdata.uspto.gov

XX Sequence 1081 BP; 310 A; 237 C; 183 G; 334 T; 0 U; 17 Other;

SQ Query Match 85.0%; Score 17; DB 6; Length 1081;

Best Local Similarity 89.5%; Pred. No. 7.3e+02;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CAAGTAAATGCAGAAACAG 19

Db 608 CAAGTAAATGCAGAAACAG 590

||||||| :|||||

RESULT 19

AAS33425

ID AAS33425 standard; DNA; 22467 BP.

XX AAS33425;

AC AAS33425;

XX 04-DEC-2001 (first entry)

DT DNA encoding human secreted protein, Seq ID No 708.

DE Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;

KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;

KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;

KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;

KW multiple sclerosis; cancer; hyperproliferative disorder; infection;

KW Gaucher's disease; neurological disease; cerebrovascular disorder;

KW thrombosis; wound healing; ds.

XX Homo sapiens.

OS WO200155326-A2.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US001347.

PF 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249284P.  
PR 17-NOV-2000; 2000US-0249285P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-451931/48.  
New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions.  
Disclosure; SEQ ID NO 708; 753pp; English.  
The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and





Mon Jan 30 07:42:54 2006

PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
EI	Rosen CA, Barash SC, Ruben SM;	
DR	WPI; 2001-451931/48.	
XX		
PT	New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions.	
PT		
PS	Disclosure; SEQ ID NO 706; 753pp; English.	
XX		
CC	The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. PCR AAS33043-AAAS33486 represent human secreted protein coding sequences, PCR	
CC		
Query Match	85.0%; Score 17; DB 4; Length 22467;	

protein identification; signal transduction pathway; metabolic pathway;  
promoter; termination sequence; ss.

KW	Arabidopsis thaliana.	PR	24-JUN-1999;	99US-0140695P.
XX	EP1033405-A2.	PR	28-JUN-1999;	99US-0140823P.
OS	06-SEP-2000.	PR	29-JUN-1999;	99US-0140991P.
PN		PR	30-JUN-1999;	99US-0141287P.
PD		PR	01-JUL-1999;	99US-0141842P.
XX		PR	01-JUL-1999;	99US-0142154P.
XX		PR	02-JUL-1999;	99US-0142055P.
XX		PR	06-JUL-1999;	99US-0142390P.
XX		PR	08-JUL-1999;	99US-0142803P.
XX		PR	09-JUL-1999;	99US-0142920P.
XX		PR	12-JUL-1999;	99US-0142977P.
XX		PR	13-JUL-1999;	99US-0143542P.
XX		PR	14-JUL-1999;	99US-0143624P.
XX		PR	15-JUL-1999;	99US-0144005P.
XX		PR	16-JUL-1999;	99US-0144085P.
XX		PR	16-JUL-1999;	99US-0144086P.
XX		PR	19-JUL-1999;	99US-0144325P.
XX		PR	19-JUL-1999;	99US-0144331P.
XX		PR	19-JUL-1999;	99US-0144332P.
XX		PR	19-JUL-1999;	99US-0144333P.
XX		PR	19-JUL-1999;	99US-0144334P.
XX		PR	19-JUL-1999;	99US-0144335P.
XX		PR	20-JUL-1999;	99US-0144352P.
XX		PR	20-JUL-1999;	99US-0144632P.
XX		PR	20-JUL-1999;	99US-0144884P.
XX		PR	21-JUL-1999;	99US-0144814P.
XX		PR	21-JUL-1999;	99US-0145086P.
XX		PR	21-JUL-1999;	99US-0145088P.
XX		PR	22-JUL-1999;	99US-0145085P.
XX		PR	22-JUL-1999;	99US-0145087P.
XX		PR	22-JUL-1999;	99US-0145089P.
XX		PR	22-JUL-1999;	99US-0145192P.
XX		PR	23-JUL-1999;	99US-0145145P.
XX		PR	23-JUL-1999;	99US-0145218P.
XX		PR	23-JUL-1999;	99US-0145224P.
XX		PR	26-JUL-1999;	99US-0145276P.
XX		PR	27-JUL-1999;	99US-0145913P.
XX		PR	27-JUL-1999;	99US-0145918P.
XX		PR	27-JUL-1999;	99US-0145919P.
XX		PR	28-JUL-1999;	99US-0145951P.
XX		PR	02-AUG-1999;	99US-0146386P.
XX		PR	02-AUG-1999;	99US-0146388P.
XX		PR	02-AUG-1999;	99US-0146389P.
XX		PR	03-AUG-1999;	99US-0147038P.
XX		PR	04-AUG-1999;	99US-0147204P.
XX		PR	04-AUG-1999;	99US-0147302P.
XX		PR	05-AUG-1999;	99US-0147192P.
XX		PR	05-AUG-1999;	99US-0147260P.
XX		PR	06-AUG-1999;	99US-0147303P.
XX		PR	06-AUG-1999;	99US-0147416P.
XX		PR	09-AUG-1999;	99US-0147493P.
XX		PR	09-AUG-1999;	99US-0147935P.
XX		PR	10-AUG-1999;	99US-0148171P.
XX		PR	11-AUG-1999;	99US-0148319P.
XX		PR	12-AUG-1999;	99US-0148341P.
XX		PR	13-AUG-1999;	99US-0148565P.
XX		PR	13-AUG-1999;	99US-0148684P.
XX		PR	16-AUG-1999;	99US-0149368P.
XX		PR	17-AUG-1999;	99US-0149175P.
XX		PR	18-AUG-1999;	99US-0149426P.
XX		PR	20-AUG-1999;	99US-0149722P.
XX		PR	20-AUG-1999;	99US-0149723P.
XX		PR	20-AUG-1999;	99US-0149829P.
XX		PR	23-AUG-1999;	99US-0149902P.
XX		PR	23-AUG-1999;	99US-0149930P.
XX		PR	25-AUG-1999;	99US-0150566P.
XX		PR	26-AUG-1999;	99US-0150884P.
XX		PR	27-AUG-1999;	99US-0151065P.
XX		PR	27-AUG-1999;	99US-0151066P.
XX		PR	27-AUG-1999;	99US-0151080P.
XX		PR	30-AUG-1999;	99US-0151303P.
XX		PR	31-AUG-1999;	99US-0151438P.
XX		PR	01-SEP-1999;	99US-0151930P.





XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
PS Disclosure; SEQ ID NO 12213; 1701pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (AB14678-AB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5626 BP; 1311 A; 1482 C; 1531 G; 1302 T; 0 U; 0 Other;  
Query Match 84.0%; Score 16.8; DB 5; Length 5626;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAACTAAATGCAGAAACAGG 20  
DB 5588 CATGTAAATGCAGAAACATG 5607  
RESULT 25  
AAQ02830  
ID AAQ02830 standard; cDNA; 9360 BP.  
XX  
AC AAQ02830;  
XX  
DT 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 29-MAY-1989 (first entry)  
XX  
DE cDNA to HIV-2 RNA.  
XX  
KW HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).  
XX  
OS Human immunodeficiency virus 2.  
XX  
Key Location/Qualifiers  
FH 544..2109  
CDS /\*tag= a  
FT /label= Gag reading frame  
FT 1830..4934  
FT /\*tag= b  
FT /label= Pol gene reading frame.  
FT 4867..5511  
FT /\*tag= c  
FT /label= Q gene reading frame  
FT 5342..5677  
FT /\*tag= d  
FT /label= X gene reading frame  
FT 5680..5994  
FT /\*tag= e  
FT /label= R gene reading frame  
FT 6148..8703  
FT /\*tag= f  
FT /label= Env gene reading frame.  
FT 8540..9304  
FT /\*tag= g  
FT

FT /label= F gene reading frame.  
XX JP01289486-A.  
PN 21-NOV-1989.  
XX  
PD 16-MAY-1988; 88JP-00119024.  
XX  
PF 16-MAY-1988; 88JP-00119024.  
XX  
PR (TOFU ) TOA NENRYO KOGYO KK.  
XX (FJRE ) FUJI REBIO KK.  
XX  
XX WPI: 1990-005177/01.  
DR P-FSDB; AAR04024, AAR04025, AAR04026, AAR04027, AAR04028, AAR04029,  
DR AAR04030.  
XX  
XX DNA indicating complement to RNA gene - of human immuno: deficiency virus  
PT type 2. used for new vaccine or diagnostic for aids virus.  
XX  
XX Claim 2; Fig 4; 12pp; Japanese.  
XX  
CC cDNA to novel HIV-2 (GH-1) has been integrated into plasmid pUC HIV-2(GH-  
CC 1). Useful for diagnosis and vaccination against the virus. Carries 7  
CC overlapping genes in various reading frames, including Gag, Pol and  
CC Env. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-  
CC 2003 to standardise OS field)  
XX  
SQ Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T; 0 U; 0 Other;  
Query Match 84.0%; Score 16.8; DB 2; Length 9360;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAACTAAATGCAGAAACAGG 20  
DB 4212 CAGGTAAATGCAGAAATAGG 4231  
RESULT 26  
AAZ89628  
ID AAZ89628 standard; DNA; 9630 BP.  
XX  
AC AAZ89628;  
XX  
DT 15-SEP-2003 (revised)  
DT 11-JUL-2000 (first entry)  
XX  
DE HIV-2 ROD DNA.  
XX  
KW Diagnosis; probe; lymphocyte; virus; immunodiagnostic; infection;  
KW antiviral; ss.  
XX  
XX Human immunodeficiency virus 2.  
XX  
XX US6054565-A.  
XX  
PD 25-APR-2000.  
XX  
PF 28-APR-1994; 94US-00234875.  
XX  
PR 03-MAR-1986; 86US-00835228.  
PR 06-OCT-1986; 86US-00916080.  
PR 21-NOV-1986; 86US-00933184.  
PR 16-JAN-1987; 87US-00003764.  
PR 11-FEB-1987; 87US-00013477.  
PR 03-SEP-1991; 91US-00752368.  
PR 20-DEC-1991; 91US-00810908.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX Montagnier L, Clavel F, Guyader M, Geutard D, Sonigo P, Alizon M;  
XX

DR WPI; 2000-328365/28.

XX Novel cloned nucleotide sequences homologous or identical to the portion

PT of genomic RNA of HIV-2 viruses useful as probes and in diagnostic tests

PT to diagnose HIV-2 infection.

XX

XX Claim 1; Col 10-23; 33pp; English.

XX

CC This invention describes a novel cloned nucleic acid (I) of a human

CC immunodeficiency virus type 2 (HIV-2). (I) is capable of being used as

CC probes in diagnostic method to obtain the immunological reagents

CC necessary to diagnose an HIV-2 infection. These sequences may be used as

CC probes in hybridization reactions with the genetic material of infected

CC patients to indicate whether the RNA of the HIV-2 virus is present in

CC these patient's lymphocytes or whether an analogous DNA is present. The

CC genetic sequence of the HIV-2 virus may be used to create the

CC polypeptides encoded by these sequences. Specifically, these polypeptides

CC may be created by expression of the cDNA obtained from bacterial, yeast

CC or animal cells. These polypeptides may be used in diagnostic tests such

CC as immunofluorescence assays, radioimmunoassays (RIA) and Western Blot

CC tests. Monoclonal antibodies to these polypeptides of fragments may be

CC created and used in immunodiagnostic tests. The polypeptides of the

CC present invention may also be used as immunogenic reagents to induce

CC protection against infection by HIV-2 viruses. The polypeptides produced

CC by recombinant-DNA techniques would function as vaccine agents. The

CC polypeptides may be used on competitive assays to test the ability of

CC various antiviral agents to determined their ability to prevent the virus

CC from fixing on its target. This sequence encodes the HIV-2 ROD genome

CC which is described in the method of the invention. (Updated on 15-SEP-

CC 2003 to standardise OS field)

XX

SQ Sequence 9630 BP; 3310 A; 1959 C; 2373 G; 1976 T; 0 U; 12 Other;

Query Match 84.0%; Score 16.8; DB 3; Length 9630;

Best Local Similarity 90.0%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||

Db 4213 CAAGTAAATGCAGAACTAGG 4232

RESULT 27

AA80890

ID AA80890 standard; cDNA; 9633 BP.

XX

AC AA80890;

XX

DT 25-MAR-2003 (revised)

DT 07-NOV-1990 (first entry)

XX

DE Sequence of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain.

XX

KW HIV vaccine; HIV strain SBL-6669-85; ss.

XX

OS Human immunodeficiency virus.

XX

Key Location/Qualifiers

FT CDS 1..9633

FT /\*tag= a

FT /note= "SQ in AAP81769"

FT CDS 2..9631

FT /\*tag= b

FT /note= "SQ in AAP81770"

FT CDS 3..9632

FT /\*tag= c

FT /note= "SQ in AAP81771"

XX

XX W08080449-A.

XX

XX 03-NOV-1988.

XX

XX 28-APR-1988; 88WO-SE000218.

XX 28-APR-1987; 87SE-00001765.

XX

XX (SBL-) SBL STATENS BAKTERI.

PA (STAT-) STATENS BAKTERIOLOGISKA LAB.

PA (STAT-) STATENS BAKTERIOLOGISKA LAB.

XX

XX Albert J, Biberfeld G, Penyo EM, Norrby E;

XX

XX WPI; 1988-322769/45.

DR P-PSDB; AAP81769, AAP81770, AAP81771.

DR

XX HIV related human retro-virus strain - used for obtaining antigens for

PT assays and vaccines and for prodn. of antibodies for assays.

PT

XX Claim 7; Fig 4; 28pp; English.

XX

CC cDNA clone (HIV-2 SBL/ISY) characterised by having the primary nucleotide

CC sequence in n80890 or a degenerate or a part thereof, esp. a clone having

CC capacity to give prodn. of infectious virus particles is claimed. HIV-2

CC SBL/ISY represents the complete genome of the virus SBL-6699 (=SBL-6689-

CC 85). The proviral DNA was obt'd. from a genomic library constructed from

CC HUT-78 cells infected with SBL-6669-85 virus using the lambda-phage

CC vector EMBL-3. SBL-6669-85 was isolated from lymphocytes of a West

CC African woman. Protection is requested for the entire genome disclosed in

CC n80890 and for parts thereof, and corresp. to various genes such as the

CC gag gene (corresp. to nucleotides 547 to 2106), the pol gene (nucleotides

CC 1827-4931) and the env gene (nucleotides 6144 to 8682), the corresp. AA

CC sequences and parts thereof and various products derived therefrom, or

CC use thereof, such as clones prep'd. by recombinant vector method, HIV test

CC devices and methods. (Updated on 25-MAR-2003 to correct PA field.)

XX

SQ Sequence 9633 BP; 3288 A; 1943 C; 2394 G; 2008 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 1; Length 9633;

Best Local Similarity 90.0%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||

Db 4209 CAAGTAAATGCAGAACTAGG 4228

RESULT 28

AA80859

ID AA80859 standard; cDNA; 9643 BP.

XX

AC AA80859;

XX

DT 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 15-NOV-1990 (first entry)

XX

DE Sequence of entire HIV-2 ROD genome.

XX

KW LAV-II ROD; ss.

XX

OS Human immunodeficiency virus; ROD.

XX

Key Location/Qualifiers

FT CDS 545..2110

FT /\*tag= a

FT /note= "AAP80801"

FT CDS 1828..4908

FT /\*tag= b

FT /note= "AAP81773"

FT CDS 4841..5485

FT /\*tag= c

FT /note= "AAP81774"

FT CDS 5316..5654

FT /\*tag= d

FT /note= "AAP81775"

FT CDS 5654..5968

```

FT FT /*tag= e
FT FT /note= "AAP81776"
FT CDS 5817..6113
FT FT /*tag= f
FT FT /note= "AAP81777"
FT CDS 6043..6153
FT FT /*tag= g
FT FT /note= "AAP81778"
FT CDS 6119..8692
FT FT /*tag= h
FT FT /note= "AAP81779"
FT CDS 8220..8372
FT FT /*tag= i
FT FT /note= "AAP81780"
FT CDS 8242..8508
FT FT /*tag= j
FT FT /note= "AAP81781"
FT CDS 8529..9299
FT FT /*tag= k
FT FT /note= "AAP81782"
FT FT
FT PN WO8805440-A.
XX XX
XX XX
XX XX 28-JUL-1988.
XX XX
XX PF 15-JAN-1989; 88WO-BF000025.
XX XX
XX PR 16-JAN-1987; 87US-00003764.
XX PR 11-FEB-1987; 87FR-00001739.
XX PR 15-APR-1987; 87FR-00005398.
XX XX
XX PA (INSP ) INST PASTEUR.
XX PA (ALIZ/) ALIZON M.
XX PA (CNRS ) CENT NAT RECH SCI.
XX XX
XX PI Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M;
XX PI Tiollais P, Chakrabar L;
XX XX
XX WPI; 1988-220290/31.
DR P-P8DB; AAP80801, AAP81773, AAP81774, AAP81775, AAP81776, AAP81777,
DR AAP81778, AAP81779, AAP81780, AAP81781, AAP81782.
XX XX
XX NEW peptide(s) with immunological properties of HIV-2 envelope protein -
PT have the structure of simian immune deficiency virus proteins, useful in
PT diagnosis and of vaccine components.
XX XX
XX PS Disclosure; Fig 1A; 86pp; French.
XX XX
XX CC The SQ in AAN80859 was deposited on 21/2/86 at the CNCM under number I-
CC 522, reference name LAV-II ROD. It is the cDNA to HIV-2 ROD genomic RNA.
CC SQ was compared with the SQ of the genome of SIV (Mac) (AAN80860) to
CC identify common regions. (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
CC correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)
XX XX
XX SQ Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 1; Length 9643;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAACAGG 20
Db 4213 CAAGTAAATGCAGAACTAGG 4232
RESULT 29
AAC81938
ID AAC81938 standard; DNA; 9663 BP.
XX AC
XX AAC81938;
XX XX
XX DT 28-FEB-2001 (first entry)

XX XX Packaging vector PCM-ROD (SD36/EM) DNA fragment.
XX DE
XX KW Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
KW cytosatic; gene therapy; transgenic; retroviral packaging;
KW gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX OS Synthetic.
XX XX WO200040741-A2.
XX PN
XX PD 13-JUL-2000.
XX XX
XX PF 06-JAN-2000; 2000WO-US000390.
XX XX
XX PR 07-JAN-1999; 99US-0115247P.
XX XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX XX
XX PI Arya SK;
XX XX
XX DR WPI; 2000-475836/41.
XX XX
XX New lentivirus transfer vector, functionally deleted for a splice donor
PT site and comprising a packaging signal and transgene operably linked to a
PT promoter, for improving encapsidation or transgene RNA and for gene
PT therapy.
XX XX
XX Example 8; Page 125-130; 143pp; English.
XX PS
XX CC This invention describes a novel transfer vector derived from a
XX CC lentivirus, functionally deleted for a splice donor site (SD), and
XX CC comprising a packaging signal and transgene operably linked to a
XX CC promoter. The products of the invention have nephrotropic,
XX CC antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
XX CC gene therapy. Encapsidation of transgene RNA is improved using the new
XX CC retroviral packaging and transfer vectors. The new transfer and packaging
XX CC vectors are used as gene delivery agents and allows transfer of a
XX CC transgene into the genome of non-dividing cells. They can be used to
XX CC create a high-efficiency packaging cell line that provides greatly
XX CC enhanced packaging of foreign DNA. Individuals suffering from a
XX CC deficiency in alpha-galactosidase expression, such as Fabry disease can
XX CC be treated by delivering the vectors to cells in vitro or in vivo.
XX CC Parkinson's disease, infectious diseases, such as acquired
XX CC immunodeficiency syndrome and cancers can be treated with the vectors.
XX CC The non-infective packaging vectors can be used to detect wild-type HIV
XX CC in biological samples using southern or northern blot assays. The
XX CC packaging of the viral RNA is maximised, without an increase in the
XX CC of the 5' SD region of the HIV-2 packaging vector results in suppressed
XX CC encapsidation of the packaging vector genomes without critical loss of
XX CC gene expression. Functional deletion of the SD site of the transfer
XX CC vector results in enhanced encapsidation of the transfer vector's genome.
XX CC HIV-2 packaging vector specifically and faithfully packages its own
XX CC optimally constructed transfer vector and gives better quality and titre
XX CC of vector than HIV-1
XX XX
XX SQ Sequence 9663 BP; 3276 A; 1952 C; 2310 G; 2074 T; 0 U; 51 Other;
Query Match 84.0%; Score 16.8; DB 3; Length 9663;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAACAGG 20
Db 4818 CAAGTAAATGCAGAACTAGG 4837
RESULT 30
ADL23697
ID ADL23697 standard; DNA; 9670 BP.
XX AC
XX ADL23697;

```



```
XX 20-MAY-2004 (first entry)
XX Human immunodeficiency virus 2 (HIV-2) ROD isolate genomic DNA.
DE ds; HIV-2 infection; LAV-II; lymphadenopathy associated virus II;
KW immunogenic; antigenic; anti-HIV; vaccine; gene.
XX Human immunodeficiency virus type 2 (ISOLATE ROD).
XX US2003235835-A1.
XX 25-DEC-2003.
XX 25-NOV-2002; 2002US-00302947.
XX 22-JAN-1986; 86FR-00000911.
XX 06-FEB-1986; 86FR-00001635.
XX 13-FEB-1986; 86FR-00001985.
XX 03-MAR-1986; 86US-00835228.
XX 18-MAR-1986; 86FR-00003881.
XX 24-MAR-1986; 86FR-00004215.
XX 06-OCT-1986; 86US-00916080.
XX 21-NOV-1986; 86US-00933184.
XX 16-JAN-1987; 87US-00003764.
XX 11-FEB-1987; 87US-00013477.
XX 03-SEP-1991; 91US-00752368.
XX 20-DEC-1991; 91US-00810908.
XX (INSP ) INST PASTEUR.
XX Alison M, Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M;
XX WPI; 2004-070575/07.
XX P-PSDB; ADL23698, ADL23699, ADL23700, ADL23701, ADL23702, ADL23703,
XX ADL23704, ADL23705, ADL23706, ADL23707, ADL23708.
XX Diagnosing HIV-2 infection by contacting genetic DNA from sample of a
XX person suspected of having HIV-2 infection with a DNA probe derived from
XX genome of HIV-2 virus and determining formation of hybridized complex.
XX Example 4; Page 6-14; 29pp; English.
XX This invention relates to diagnosing an HIV-2 (or LAV-II, lymphadenopathy
XX associated virus II) infection using an HIV-2 hybridisation DNA probe.
XX Specifically, it provides an in vitro kit capable of diagnosing an
XX infection caused by the HIV-2 virus that comprises contacting a
XX biological fluid with an immunogenic or antigenic peptide (or conjugate
XX thereof) and detecting an antigen-antibody complex by physical or
XX chemical methods. The present invention describes a detection step
XX performed by radioimmunoassays (RIA), radioimmunoprecipitation assays
XX (RIPA), immunofluorescence assays (IFA) or enzyme-linked immunosorbent
XX assays (ELISA). Accordingly, these anti-HIV compositions are useful for
XX developing vaccines to induce production of antibodies against infection
XX by HIV-2 viruses. This polynucleotide sequence is the HIV-2 ROD isolate
XX genomic DNA sequence of the invention.
XX
XX Sequence 9670 BP; 3310 A; 1992 C; 2382 G; 1983 T; 0 U; 3 Other;
Query Match 84.0%; Score 16.8; DB 12; Length 9670;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAACAGG 20
Db 4214 CAAGTAAATGCAGAACTAGG 4233
RESULT 31
AAQ20616
ID AAQ20616 standard; DNA; 9672 BP.
XX
XX AAQ20616;
AC
```

```
XX 09-SEP-2004 (revised)
XX 24-OCT-2003 (revised)
XX 25-MAR-2003 (revised)
XX 19-MAY-1992 (first entry)
XX ROD HIV-2 isolate complete genome.
XX Human immunodeficiency virus; AIDS; ss.
XX Human immunodeficiency virus 2.
XX Unidentified.
XX Key Location/Qualifiers
XX LTR 1..173
XX /tag= a
XX /note= "LTR:R 1"
XX misc_feature 174..299
XX /tag= b
XX /note= "U5"
XX CDS 546..2111
XX /tag= c
XX /note= "GAG gene"
XX misc_feature 951..1640
XX /tag= d
XX /note= "GAG p 26"
XX misc_feature 1701..2111
XX /tag= e
XX /note= "GAG p 12"
XX misc_feature 1829..4936
XX /tag= f
XX /note= "polymerase"
XX misc_feature 4869..5513
XX /tag= g
XX /note= "O protein"
XX misc_feature 5344..5679
XX /tag= h
XX /note= "X protein"
XX misc_feature 5682..5996
XX /tag= i
XX /note= "Y protein"
XX exon 5845..6140
XX /tag= j
XX /note= "TAT gene exon 1"
XX exon 6071..6140
XX /tag= k
XX /note= "ART gene exon 1"
XX misc_feature 6147..8720
XX /tag= l
XX /note= "Env protein"
XX exon 8307..8536
XX /tag= m
XX /note= "ART gene exon 2"
XX misc_feature 8557..9324
XX /tag= n
XX /note= "F protein"
XX misc_feature 8942..9497
XX /tag= o
XX /note= "U3"
XX LTR 9498..9671
XX /tag= p
XX /note= "LTR:R 2"
XX
XX USS079342-A.
XX 07-JAN-1992.
XX
XX 11-FEB-1987; 87US-00013477.
XX
XX 03-MAR-1986; 86US-00835228.
XX 06-OCT-1986; 86US-00916080.
XX 21-NOV-1986; 86US-00933184.
XX 16-JAN-1987; 87US-00003764.
```

```

XX PA (INSP ) INST PASTEUR.
XX PI Alison M, Montagnier L, Geutard D, Clavel F, Sonigo P, Guyader M;
XX PS WPI; 1992-041067/05.
XX DR P-PSDB; AAR20596, AAR20597, AAR20598, AAR20599, AAR20600, AAR20601,
XX DR AAR20602, AAR20603, AAR20604, AAR20605, AAR20606, AAR20607.
XX PT Peptide(s) corresp. to HIV-2 aminoacid sequences - used in diagnosis in
XX PT vaccines and in prodn. of antibodies for diagnosis.
XX PS Disclosure; Page 12; 30pp; English.
XX CC The sequence is that of the complete genomic sequence of the ROD HIV-2
XX CC isolate. Its genes and their major expression products are indicated in
XX CC the features. It was isolated from a CEM cell line infected with the ROD
XX CC isolate. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 24-OCT
XX CC -2003 to standardise OS field)
XX CC Revised record issued on 09-SEP-2004 : Correction to Feature Table Key
XX CC
XX SQ Sequence 9672 BP; 3314 A; 1981 C; 2394 G; 1983 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 2; Length 9672;
XX Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 CAAGTAAATGCAGAACACAGG 20
DB 4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 32
AAN92119
ID AAN92119 standard; DNA; 9693 BP.
XX AC AAN92119;
XX DT 25-MAR-2003 (revised)
XX DT 17-DEC-2001 (revised)
XX DT 06-APR-1990 (first entry)
XX DE Sequence of clone HIV-2 SBL/ISY.
XX KW HIV-2; proviral clone HIV-2 SBL/ISY.
XX OS Human immunodeficiency virus 2.
XX FH Key Location/Qualifiers
XX FT CDS 1..9693
XX FT /*tag= a
XX FT CDS 2..9691
XX FT /*tag= b
XX FT CDS 3..9692
XX FT /*tag= c
XX PN USN7331212-N.
XX PD 29-AUG-1989.
XX PF 31-MAR-1989; 89US-00331212.
XX PR 31-MAR-1989; 89US-00331212.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX PA (USDC ) US SEC OF COMMERCE.
XX PA (USDC ) US SEC OF COMMERCE.
XX PI Franchini G, Wongstaal F, Gallo R;
XX WPI; 1989-339698/46.
XX DR P-PSDB; AAP93283. AAP93284. AAP93285.

```

---

```

XX PT Complete human immuno-deficiency type 2 pro-viral clone - used to
XX PT generate animal model for function studies of HIV genes in vivo.
XX PS Disclosure; Fig 5; 43pp; English.
XX CC HIV-2 SBL/ISY is a proviral clone of HIV-2. It contains a complete
XX CC provirus including the cellular flanking sequences. The clone is
XX CC infectious and cytopathic for some human T cell lines, induces syncytia
XX CC and infects cell lines in vitro. The clone can be used to generate an
XX CC animal model for functional studies of HIV genes in vivo, and to develop
XX CC experimental approaches to prevent retroviral infection in man. It is
XX CC constructed using DNA from neoplastic human cell line HUT78 freshly
XX CC infected with the HIV-2 SBL6669 viral isolate. It can be translated in
XX CC all 3 reading frames. (Note: Revised entry submitted to correct the
XX CC patent number format of US Government-owned NTIS applications to prevent
XX CC clashes with ongoing US granted patent numbers. For further information
XX CC please visit the Derwent web site at
XX CC www.derwent.com/dwpi/updates/ntis.us.html.) (Updated on 25-MAR-2003 to
XX CC correct PR field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated
XX CC on 25-MAR-2003 to correct PI field.)
XX SQ Sequence 9693 BP; 3315 A; 1947 C; 2412 G; 2019 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 1; Length 9693;
XX Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 CAAGTAAATGCAGAACACAGG 20
DB 4269 CAAGTAAATGCAGAACTAGG 4288

RESULT 33
AAC81937
ID AAC81937 standard; DNA; 9726 BP.
XX AC AAC81937;
XX DT 28-FEB-2001 (first entry)
XX DE Packaging vector PROD(SD36/EM) DNA fragment.
XX KW Encapsulation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
XX KW cytostatic; gene therapy; transgenic; retroviral packaging;
XX KW gene delivery; Parkinson's disease; infectious diseases; cancer; de.
XX OS Synthetic.
XX PN WO200040741-A2.
XX PD 13-JUL-2000.
XX PF 06-JAN-2000; 2000WO-US000390.
XX PR 07-JAN-1999; 99US-0115247P.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Arya SK;
XX WPI; 2000-475836/41.
XX DR New lentivirus transfer vector, functionally deleted for a splice donor
XX PT site and comprising a packaging signal and transgene operably linked to a
XX PT promoter, for improving encapsidation or transgene RNA and for gene
XX PT therapy.
XX PS Example 6; Page 119-124; 143pp; English.
XX CC This invention describes a novel transfer vector derived from a
XX CC lentivirus, functionally deleted for a splice donor site (SD), and
XX CC comprising a packaging signal and transgene operably linked to a

```



```
XX AC ADC87444;
XX DT 01-JAN-2004 (first entry)
XX DE Human GPCR gene SEQ ID NO:1897.
XX KW ds; gene; human; GPCR;
XX KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX OS Homo sapiens.
XX PN EP1270724-A2.
XX PD 02-JAN-2003.
XX PF 18-JUN-2002; 2002EP-00013517.
XX PR 18-JUN-2001; 2001JP-00246789.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX DR WPI; 2003-315783/31.
XX DR P-PSDB; ADC87445.
XX PT New polynucleotide, useful for preparing a composition for treating a
XX PT patient in need of increased or suppressed activity or expression of the
XX PT guanosine triphosphate-binding protein coupled receptor.
XX PS Claim 1; SEQ ID NO 1897; 28pp; English.
XX CC The invention relates to a novel polynucleotide encoding a guanosine
XX CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX CC the invention may have a use in gene therapy. The polynucleotide and
XX CC polypeptide are useful for preparing a composition for treating a patient
XX CC in need of increased or suppressed activity or expression of the
XX CC guanosine triphosphate-binding protein coupled receptor. The
XX CC polynucleotide sequences shown in ADC85548-ADC87615 encode GPCR's of the
XX CC invention.
XX SQ Sequence 16270 BP; 4172 A; 3573 C; 4062 G; 4463 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 10; Length 16270;
XX Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAATGCGAAGAACAGG 20
XX DB 6075 CAAGTAATGCGAAGAACAGG 6056
XX
XX RESULT 37
XX ABD33401
XX ID ABD33401 standard; DNA; 41079 BP.
XX AC ABD33401;
XX XX
XX DT 18-NOV-2004 (first entry)
XX DE Human cancer-associated (CA) gene HD07-075.
XX KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX KW ds; cancer; cytostatic.
XX OS Homo sapiens.
XX PN WO2004058146-A2.
XX PD 15-JUL-2004.
XX
XX PF 15-DEC-2003; 2003WO-US040081.
XX PR 17-DEC-2002; 2002US-00322281.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX XX
XX DR WPI; 2004-499109/47.
XX PT Novel human cancer associated protein encoded within open reading frame
XX PT of cancer associated gene, useful as targets for diagnosing cancer.
XX PS Claim 16; SEQ ID NO 516; 182pp; English.
XX CC The invention relates to cancer-associated proteins (CAP) and the cancer-
XX CC associated (CA) nucleic acids encoding them. The invention also relates
XX CC to a method for treating cancers involving administering to a patient an
XX CC inhibitor of CAP, and a method of screening for anticancer activity in a
XX CC potential drug involving providing a cell that expresses a CA gene,
XX CC contacting a tissue sample derived from a cancer cell with an anticancer
XX CC drug candidate and monitoring the effect of the anticancer drug candidate
XX CC on expression of the CA gene. The CAP proteins are useful for detecting
XX CC cancer associated with expression of a CAP protein in a test cell sample
XX CC and for screening for a bioactive agent capable of modulating the
XX CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX CC cancer, involving determining the expression of a CA nucleic acid in a
XX CC tissue. This sequence represents a human CA gene of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 41079 BP; 8798 A; 10965 C; 11900 G; 9118 T; 0 U; 298 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 13; Length 41079;
XX Best Local Similarity 90.0%; Pred. No. 1.2e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAATGCGAAGAACAGG 20
XX DB 2956 CATGTAATGCGAAGAACATG 2975
XX
XX RESULT 38
XX ADZ13504
XX ID ADZ13504 standard; DNA; 41104 BP.
XX AC ADZ13504;
XX XX
XX DT 16-JUN-2005 (first entry)
XX DE Human cancer-associated genomic DNA #87.
XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX KW cytostatic; gene; ds.
XX OS Homo sapiens.
XX PN WO2005031001-A2.
XX PD 07-APR-2005.
XX PF 23-SEP-2004; 2004WO-US031617.
XX PR 23-SEP-2003; 2003US-00669920.
XX PA (CHIR ) CHIRON CORP.
XX XX
XX PI Morris DW, Malandro MS;
XX DR WPI; 2005-273395/28.
XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,
```



XX 21-AUG-1997.  
PD  
XX  
XX  
PF 19-FEB-1997; 97WO-US002318.  
XX  
XX  
PR 20-FEB-1996; 96US-0011888P.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
PA  
XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
XX  
XX WPI; 1997-424969/39.  
DR P-PSDB; AAW27790.  
XX  
XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used to  
PT isolate antimicrobial compounds, and in vaccines against S. aureus  
PT infection.  
XX  
XX Claim 9; Page 657; 989pp; English.  
PS  
XX The present sequence encodes a Staphylococcus aureus protein of unknown  
CC function. The present sequence was isolated from a library of clones of  
CC S. aureus WCUH 29 in Escherichia coli. The DNA sequences can be used in  
CC the construction of ribozymes and antisense sequences to control the  
CC expression of Staphylococcal genes. The DNA sequences is also useful as a  
CC source of regulatory elements for the control of bacterial gene  
CC expression. The encoded protein may be used to produce vaccines to enable  
CC a host to produce specific antibodies with antibacterial action. These  
CC vaccines and antibodies would protect a host against invasion by S.  
CC aureus, and conditions relating to Staphylococcal infection, e.g.  
CC Staphylococcal food poisoning, scaled skin syndrome, and toxic shock  
CC syndrome  
XX  
SQ Sequence 453 BP; 152 A; 72 C; 97 G; 130 T; 0 U; 2 Other;  
SQ  
Query Match 82.0%; Score 16.4; DB 2; Length 453;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACA 18  
DB 312 CAAGTAAATGCAGAAATCA 295  
AC  
ADQ55786;  
ADQ55786/c  
ID ADQ55786 standard; DNA; 469 BP.  
XX  
AC  
ADQ55786;  
ADQ55786;  
DT 21-OCT-2004 (first entry)  
XX  
DE Novel canine microarray-related DNA sequence SeqID7088.  
XX  
XX canine microarray; drug screening; toxicity assay;  
KW environmental pollutant; cellular response; gene expression profile;  
KW toxic response; liver necrosis; fatty liver disease;  
KW protein adduct formation; hepatitis; dog; ds.  
XX  
XX Canis familiaris.  
OS  
XX WO2004063324-A2.  
PN  
XX 29-JUL-2004.  
PD  
XX 05-MAY-2003; 2003WO-US013853.  
PF  
XX 03-MAY-2002; 2002US-0377240P.  
PR  
XX (GENE-) GENE LOGIC INC.  
PA (PFIZ ) PFIZER PROD INC.  
XX

PI Diggans JC, Porter M, Wei T;  
XX  
XX WPI; 2004-561890/54.  
XX  
XX New isolated nucleic acid molecule, useful for drug screening and  
PT toxicity assays or for assessing the impact, including toxicity, of a  
PT compound, pharmaceutical agent or environmental pollutant on a cell or  
PT living organism.  
XX  
XX Claim 1; SEQ ID NO 7088; 41pp; English.  
PS  
XX This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and  
CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers  
CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver  
CC disease, protein adduct formation or hepatitis), those of the kidney,  
CC heart, brain or testes, or other pathologies associated with at least one  
CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention.  
XX  
SQ Sequence 469 BP; 135 A; 69 C; 62 G; 160 T; 0 U; 43 Other;  
SQ  
Query Match 82.0%; Score 16.4; DB 13; Length 469;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACA 18  
DB 304 CAAGTAAATGCAGAAAAA 287  
AC  
ABX08116;  
ABX08116/c  
ID ABX08116 standard; DNA; 507 BP.  
XX  
AC  
ABX08116;  
ABX08116;  
DT 27-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain coding region #2404.  
XX  
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
KW auditory; respiratory; gene therapy; vaccine.  
XX  
XX Streptococcus pneumoniae; type 4 strain.  
OS  
XX WO200277021-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 27-MAR-2002; 2002WO-IB002163.  
PF

XX 27-MAR-2001; 2001GB-00007658.  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX Maignani V, Tettelin H, Fraser C;  
XX WPI; 2003-040579/03.  
XX P-PSDB; ABU02823.  
XX  
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
XX useful as medicaments for treating or preventing a disease or infection  
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
XX ear infection.  
XX  
XX Claim 6; SEQ ID NO 4807; 56pp; English.  
XX  
XX The invention relates to a protein comprising or having at least 50%  
XX identity to any of the 2469 amino acid sequences, identified in the  
XX specification (available on a computer readable format), or its fragment,  
XX expressed from 2469 of 2489 identified DNA coding regions from the  
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
XX AB556454. Also included are an antibody which binds one of the proteins,  
XX treating a patient by administering the protein, DNA or antibody (in a  
XX composition), a kit comprising first and second primers, which are the  
XX nucleic acid cited above or fragments between nucleotides 8-100 of a  
XX sequence not defined in the specification, for amplifying a target  
XX sequence contained within a Streptococcus nucleic acid sequence, where  
XX the first primer is substantially complementary to the target sequence  
XX and the second primer is substantially complementary to the complement of  
XX the target sequence, and where the parts of the primers having  
XX substantial complementarity define the termini of the target sequence to  
XX be amplified, assay comprising contacting a test compound with the  
XX protein, and determining whether the test compound binds to the protein  
XX and a Streptococcus pneumoniae bacterium, where one or more genes  
XX encoding the proteins has been rendered inactive. The proteins, nucleic  
XX acid molecules, antibody and compositions are useful as medicaments for  
XX treating or preventing a disease or infection due to streptococcus  
XX bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
XX media or ear infection. They are also useful in developing vaccines,  
XX diagnostics and antibiotics. The methods are useful for identifying  
XX immunodominant proteins. The present sequence is one of the 2489  
XX identified coding region from the genomic sequence. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
XX standardise OS field)  
XX  
XX Sequence 507 BP; 149 A; 64 C; 85 G; 209 T; 0 U; 0 Other;  
XX  
XX Query Match 82.0%; Score 16.4; DB 10; Length 507;  
XX Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 2 AAGTAAATGCAGAAACAG 19  
XX ||||| ||||| ||||| |||||  
XX Db 501 AAGTAAATGCAGTAAACAG 484  
XX  
XX RESULT 43  
XX ADM91806/c  
XX ID ADM91806 standard; DNA; 507 BP.  
XX XX  
XX AC ADM91806;  
XX  
XX DT 03-JUN-2004 (first entry)  
XX  
XX DE S pneumoniae antigenic protein-encoding gene sequence SeqID3.  
XX KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
XX KW antigenic; gene; ds.  
XX

OS Streptococcus pneumoniae.  
XX WO2004020609-A2.  
XX PD 11-MAR-2004.  
XX  
XX PF 02-SEP-2003; 2003WO-US027401.  
XX PR 30-AUG-2002; 2002US-0407082P.  
XX XX (TUFT ) UNIV TUFTS.  
XX PI Camilli A, Hava DL;  
XX DR WPI; 2004-239189/22.  
XX DR P-PSDB; ADM92043.  
XX  
XX PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
XX diagnosing, treating and preventing active infections of Streptococcus  
XX pneumoniae.  
XX PS Claim 1; SEQ ID NO 3; 123pp; English.  
XX  
XX CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
XX acid molecules and the antigenic polypeptides encoded by them. The  
XX invention may be useful for the production of compounds with an  
XX antibacterial activity or for gene therapy. The nucleic acid molecules,  
XX compositions and methods disclosed are useful for treating Streptococcus  
XX pneumoniae infection. The present sequence is that of an S pneumoniae  
XX gene of the invention.  
XX  
XX SQ Sequence 507 BP; 149 A; 64 C; 85 G; 209 T; 0 U; 0 Other;  
XX  
XX Query Match 82.0%; Score 16.4; DB 12; Length 507;  
XX Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 2 AAGTAAATGCAGAAACAG 19  
XX ||||| ||||| ||||| |||||  
XX Db 501 AAGTAAATGCAGTAAACAG 484  
XX  
XX RESULT 44  
XX ACH75526  
XX ID ACH75526 standard; DNA; 512 BP.  
XX XX  
XX AC ACH75526;  
XX  
XX DT 29-JUL-2004 (first entry)  
XX  
XX DE Human genome derived single exon probe #8721.  
XX  
XX KW Human; probe; ss; gene expression; single exon probe; microarray;  
XX KW alternative splicing event; genomic alteration.  
XX  
XX OS Homo sapiens.  
XX  
XX PN US2003194704-A1.  
XX  
XX PD 16-OCT-2003.  
XX  
XX PF 03-APR-2002; 2002US-00029386.  
XX  
XX PR 03-APR-2002; 2002US-00029386.  
XX  
XX PA (PENN/) PENN S G.  
XX PA (RANK/) RANK D R.  
XX PA (HANZ/) HANZEL D K.  
XX  
XX PI Penn SG, Rank DR, Hanzel DK;  
XX DR WPI; 2004-119264/12.  
XX





XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
 KW enzymatic assay; antibiotic target; gene; ds.  
 XX Staphylococcus aureus.  
 OS WO200294868-A2.  
 PN 28-NOV-2002.  
 PD 27-MAR-2002; 2002WO-IB002637.  
 PF 27-MAR-2001; 2001GB-00007661.  
 XX (CHIR-) CHIRON SPA.  
 XX Masignani V, Mora M, Scarselli M;  
 PI WPI; 2003-120786/11.  
 DR P-PSDB; ABM72316.  
 XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
 PT preventing Staphylococcal infection, specifically an infection caused by  
 PT S. aureus, e.g. sepsis.  
 XX Claim 6; SEQ ID NO 3111; 49pp; English.  
 PS The invention relates to novel genes and encoded proteins from  
 CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
 CC nucleic acid encoding the protein, or an antibody to the protein, is  
 CC useful as a pharmaceutical, particularly as a vaccine for treating or  
 CC preventing infection due to Staphylococcus bacteria, specifically an  
 CC infection caused by S. aureus. The composition is particularly useful for  
 CC treating or preventing sepsis in a patient. The composition can also be  
 CC used for diagnostics. The protein is also used in an assay for enzymatic  
 CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel S. aureus genes of the invention  
 XX  
 SQ Sequence 711 BP; 241 A; 117 C; 141 G; 212 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 8; Length 711;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACA 18  
 |||||  
 Db 137 CAAGTAAATGCAGAAATCA 120  
 RESULT 47  
 ABN93331/c  
 ID ABN93331 standard; DNA; 714 BP.  
 AC ABN93331;  
 XX 24-JUL-2002 (first entry)  
 DT Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2794.  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 DE antibacterial; gene therapy; gene; ds.  
 KW Staphylococcus epidermidis.  
 XX US6380370-B1.  
 XX 30-APR-2002.  
 PD 13-AUG-1998; 98US-00134001.  
 PF 14-AUG-1997; 97US-0055779P.  
 XX 08-NOV-1997; 97US-0064964P.  
 PR

PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 DR P-PSDB; ABP40786.  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX Disclosure; SEQ ID NO 2794; 267pp; English.  
 PS ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX Sequence 714 BP; 247 A; 108 C; 137 G; 222 T; 0 U; 0 Other;  
 SQ Query Match 82.0%; Score 16.4; DB 6; Length 714;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACA 18  
 |||||  
 Db 137 CAAGTAAATGCAGAAATCA 120  
 RESULT 48  
 ADS04339/c  
 ID ADS04339 standard; DNA; 714 BP.  
 AC ADS04339;  
 XX 04-NOV-2004 (first entry)  
 DT Staphylococcus epidermis polynucleotide seqid 3634.  
 DE antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;  
 KW recombinant expression vector; infection; computer readable medium;  
 KW computer based system; gene; ds.  
 XX Staphylococcus epidermidis.  
 OS US2004147734-A1.  
 XX 29-JUL-2004.  
 PD 01-DEC-2003; 2003US-00724972.  
 PF 08-NOV-1997; 97US-0064964P.  
 PR 13-AUG-1998; 98US-00134001.  
 XX 29-NOV-1999; 99US-00450969.  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 XX Doucette-Stamm L, Bush D;  
 PI WPI; 2004-580138/56.  
 DR P-PSDB; ADS08111.  
 XX New isolated polypeptide and encoding nucleic acid derived from  
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
 PT treating an S. epidermidis bacterial infection.  
 XX Claim 5; SEQ ID NO 3634; 741pp; English.  
 PS

XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an *Staphylococcus epidermidis* polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an *S. epidermidis* polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for *S. epidermidis* infection; a recombinant or substantially  
 CC pure preparation of an *S. epidermidis* polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*  
 CC infection; detecting the presence of a *Staphylococcus* nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the *Staphylococcus* genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the *Staphylococcus* plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the *Staphylococcus*  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the *Staphylococcus* genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an *Staphylococcus epidermidis* bacterial  
 CC infection. This sequence encodes a *S. epidermidis* protein of the invention.

XX Sequence 714 BP; 247 A; 108 C; 137 G; 222 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 13; Length 714;

Best Local Similarity 94.4%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAAGCA 18

Db 137 CAAGTAATGCGAAGTCA 120

RESULT 49

ACA24380

ID ACA24380 standard; DNA; 762 BP.

AC ACA24380;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #6037.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.

OS Bacteroides fragilis.

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU20510.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 XX isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 12250; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC the gene product or that has an activity against a biological pathway  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 762 BP; 274 A; 147 C; 180 G; 161 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 8; Length 762;

Best Local Similarity 94.4%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCGAAGCAACAG 19

Db 296 AAGTAATGCGAAGCAACAG 313

RESULT 50

AAV75056

ID AAV75056 standard; DNA; 924 BP.

AC AAV75056;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #745.

XX Computer readable medium; vaccine; *S. aureus* infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

EH Key Location/Qualifiers

FT misc\_feature 481..540

FT /\*tag= a

FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They are  
 FT included to maintain the nucleotide numbering given in  
 FT the specification for this DNA sequence"

XX EP786519-A2.  
 PN 30-JUL-1997.  
 PD 07-JAN-1997; 97EP-00100117.  
 XX 05-JAN-1996; 96US-0009861P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Kunsch CA, Choi GH, Baraash SC, Dillon PJ, Fannon MR, Rosen CA;  
 PI WPI; 1997-374922/35.  
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 PT stored on computer readable medium and used in the production of anti-  
 PT *S.aureus* vaccines.  
 XX Claim 1; Page 1641-1642; 3271pp; English.  
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the *S.aureus* DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against *S.aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the *S.aureus* DNA sequences contained on the computer  
 CC readable medium  
 XX  
 XX Sequence 924 BP; 258 A; 168 C; 148 G; 289 T; 0 U; 61 Other;  
 SQ  
 Query Match 82.0%; Score 16.4; DB 2; Length 924;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAAGTAAATGCAGAAACA 18  
 |||||  
 DB 273 CAAAGTAAATGCAGAAACA 290  
 |||||  
 RESULT 51  
 AAT43303  
 ID AAT43303 standard; DNA; 1035 BP.  
 XX AAT43303;  
 AC  
 XX 10-FEB-1997 (first entry)  
 DT  
 XX Vmp7 soluble variant coding sequence.  
 DE  
 XX OspA; OspB; outer surface protein A; Borrelia; variable major protein 7;  
 KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;  
 KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KW ds.  
 XX Borrelia hermslii.  
 OS  
 XX US5571718-A.  
 PN  
 XX 05-NOV-1996.  
 PD  
 XX 08-SEP-1992; 92US-00941523.  
 PF  
 XX 21-DEC-1990; 90US-00632072.  
 PR  
 XX (ASUY-) ASSOC UNIVERSITIES INC.  
 PA Barbour AG, Dunn JJ;  
 PI

PF 08-SEP-1992; 92US-00941523.  
 XX 21-DEC-1990; 90US-00632072.  
 PR (ASUY-) ASSOC UNIVERSITIES INC.  
 XX Barbour AG, Dunn JJ;  
 PI WPI; 1996-505409/50.  
 XX P-PSDB; AAW08097.  
 XX Soluble recombinant forms of Borrelia lipo:proteins - useful for vaccine  
 PT prodn. for treatment of Lyme disease.  
 PT Claim 4; Col 43-46; 49pp; English.  
 XX AAT43301-743303 represent coding sequences for soluble recombinant  
 CC Borrelia proteins of the invention. This sequence represents the coding  
 CC sequence for the soluble recombinant variant of the Borrelia hermslii  
 CC variable major protein 7 (Vmp7). Borrelia spirochetes are responsible for  
 CC a variety of human disorders including Lyme borreliosis, and relapsing  
 CC fevers. The spirochete is transmitted to humans and animals through the  
 CC bite of a tick, and can cause serious dermatological, arthritic,  
 CC neurological and other pathogenic disorders in an infected host. This  
 CC sequence is used to create recombinant host cells, and the encoded Vmp7  
 CC protein can be isolated from the cytosol of one of these cells without  
 CC the use of detergent. The encoded recombinant proteins can be used as  
 CC antigens for the production of vaccines against Lyme disease. The  
 CC recombinant proteins can also be used in immunoassays and other  
 CC diagnostic screening methods to detect the presence of antibodies against  
 CC Borrelia lipoproteins in the sera of infected patients  
 XX  
 XX Sequence 1035 BP; 397 A; 141 C; 274 G; 223 T; 0 U; 0 Other;  
 SQ  
 Query Match 82.0%; Score 16.4; DB 2; Length 1035;  
 Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 AGTAAATGCAGAAACAGG 20  
 |||||  
 DB 246 AGTAAAGCAGAAACAGG 263  
 |||||  
 RESULT 52  
 AAT43316  
 ID AAT43316 standard; DNA; 1110 BP.  
 XX AAT43316;  
 AC  
 XX 10-FEB-1997 (first entry)  
 DT  
 XX Vmp7 variant #1 coding sequence.  
 DE  
 XX OspA; OspB; outer surface protein A; Borrelia; variable major protein 7;  
 KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;  
 KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KW ds.  
 XX Synthetic.  
 OS  
 XX US5571718-A.  
 PN  
 XX 05-NOV-1996.  
 PD  
 XX 08-SEP-1992; 92US-00941523.  
 PF  
 XX 21-DEC-1990; 90US-00632072.  
 PR  
 XX (ASUY-) ASSOC UNIVERSITIES INC.  
 PA Barbour AG, Dunn JJ;  
 PI

DR WPI; 1996-505409/50.

XX Soluble recombinant forms of *Borrelia lipo*:proteins - useful for vaccine

PT prodn. for treatment of Lyme disease.

XX Example 9; Col 41-42; 49pp; English.

XX This sequence represents the coding sequence for a soluble recombinant

CC *Borrelia* variable major protein 7 (Osp7) variant protein of the

CC invention. *Borrelia* spirochetes are responsible for a variety of human

CC disorders including Lyme borreliosis, and relapsing fevers. The

CC spirochete is transmitted to humans and animals through the bite of a

CC tick, and can cause serious dermatological, arthritic, neurological and

CC other pathogenic disorders in an infected host. This sequence is used to

CC create recombinant host cells, and the encoded vmp7 protein can be

CC isolated from the cytosol of one of these cells without the use of

CC detergent. The encoded recombinant proteins can be used as antigens for

CC the production of vaccines against Lyme disease. The recombinant proteins

CC can also be used in immunoassays and other diagnostic screening methods

CC to detect the presence of antibodies against *Borrelia* lipoproteins in the

CC sera of infected patients

XX

SQ Sequence 1110 BP; 432 A; 143 C; 286 G; 249 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 2; Length 1110;

Best Local Similarity 94.4%; Pred. No. 1.4e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAGG 20

DB 321 AGTAAACGACGAAACAGG 338

RESULT 53

AAA26350/C

ID AAA26350 standard; cDNA; 1196 BP.

XX AAA26350;

AC

XX 29-JUN-2000 (first entry)

DT

XX Human secreted protein gene 5 SEQ ID NO:15.

DE

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

KW antiHIV; antinflammatory; nontropic; neuroprotective; antiallergic;

KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;

KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;

KW immune disease; inflammation; blood disorder; tumour; ss.

OS Homo sapiens.

XX

XX WO200006698-A1.

FN

XX 10-FEB-2000.

PD

XX 29-JUL-1999; 99WO-US017130.

FF

XX 30-JUL-1998; 98US-0094657P.

PR

XX 05-AUG-1998; 98US-0095486P.

PR

XX 06-AUG-1998; 98US-0095454P.

PR

XX 06-AUG-1998; 98US-0095455P.

PR

XX 12-AUG-1998; 98US-0096319P.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

XX Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

XX

XX WPI; 2000-195282/17.

DR

XX P-PSDB; AAY91455.

XX

XX New isolated human genes and the secreted polypeptides they encode,

PT

useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.

Claim 1; Page 374; 634pp; English.

The polynucleotide sequences given in AAA26346 to AAA26458 encode the human secreted proteins given in AAY91451 to AAY91691. The human secreted proteins can have activities based on the tissues and cells they are expressed in. Examples of the activities are: cytostatic; immunosuppressive; antiHIV; antinflammatory; nontropic; neuroprotective; antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic; cardiant. The polynucleotides and their corresponding secreted proteins are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Also pathological conditions can be diagnosed by determining the amount of the proteins in a sample or by determining the presence of mutations in the polynucleotides. Specific uses are described for each of the polynucleotides, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, diseases of the immune system, autoimmune diseases, hepatic and renal disease, inflammation, allergies, Alzheimer's and behavioural disorders, schizophrenia, osteoporosis, arthritis, infections, AIDS, spinal cord injuries, transplant rejection, diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders, reproductive disorders, gastrointestinal disorders, respiratory disorders and metabolic disorders. The proteins or polynucleotides can also be used as food additives or preservatives. The proteins are also useful for identifying their binding partners. AAA26337 to AAA26345 and AAY91450 are sequences used in the exemplification of the present invention

Sequence 1196 BP; 333 A; 171 C; 192 G; 500 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 3; Length 1196;

Best Local Similarity 94.4%; Pred. No. 1.4e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACA 18

DB 311 CAATTAATGCAGAAACA 294

RESULT 54

ADL71411/C

ID ADL71411 standard; cDNA; 1196 BP.

XX ADL71411;

AC

XX 20-MAY-2004 (first entry)

DT

XX Novel human secreted protein cDNA seqid 15.

DE

XX antinflammatory; neuroprotective; nontropic; antiparkinsonian;

KW anticonvulsant; antilipaeamic; CNS; gynaecological; antiarthritic;

KW antiasthmatic; anti-HIV; virucide; endocrine; cytostatic;

KW immunosuppressive; antiallergic; cardiovascular; respiratory;

KW dermatological; antimicrobial; gastrointestinal; gene therapy;

KW neurodegenerative disease; behavioural disorder; inflammatory condition;

KW hyperproliferative disorder; Alzheimer's disease; Parkinson's disease;

KW Huntington's disease; metabolic disorder; Tay-Sach's disease;

KW Leash-Nyhan syndrome; reproductive disorder; immunological disorder;

KW arthritis; asthma; AIDS; endocrine disorder; immune disorder;

KW Hodgkin's lymphoma; haematopoietic disorder; muscular disorder;

KW leukaemia; autoimmune disorder; allergy; cancer; cardiovascular disorder;

KW respiratory disorder; pulmonary disorder; connective tissue disorder;

KW skin disorder; CNS disorder; congenital disorder; infectious disorder;

KW gastrointestinal disorder; human; secreted protein; gene; ss.

OS Homo sapiens.

XX

XX US2004034196-A1.

FN

XX 19-FEB-2004.

PD



DB 334 AAGTAGATGCAGAAACAG 351  
||||| |||||||  
RESULT 56  
AB054797  
ID AB054797 standard; cDNA; 1808 BP.  
XX AC AB054797;  
XX DT 22-AUG-2002 (first entry)  
XX DE Human ovarian antigen HTHD776 cDNA, SEQ ID NO:677.  
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200200677-A1.  
XX PD 03-JAN-2002.  
XX PF 07-JUN-2001; 2001WO-US018569.  
XX PR 07-JUN-2000; 2000US-0209467P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Birse CE, Rosen CA;  
XX DR WPI; 2002-147878/19.  
XX DR P-PSDB; ABP41720.  
XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX PS Claim 1; SEQ ID NO 677; 2922pp; English.  
XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies

CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1808 BP; 430 A; 464 C; 484 G; 427 T; 0 U; 3 Other;  
Query Match 82.0%; Score 16.4; DB 6; Length 1808;  
Best Local Similarity 85.0%; Pred. No. 1.4e+03;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 20  
DB 1770 CATGTAATGCGAAACATG 1789  
RESULT 57  
ABA90791  
ID ABA90791 standard; DNA; 2014 BP.  
XX AC ABA90791;  
XX DT 14-FEB-2002 (first entry)  
XX DE Bacillus anthracis DNA fingerprinting PCR primer SEQ ID NO 22.  
XX KW Bacillus anthracis; MLVA primer; VNTR; DNA finger printing; anthrax;  
KW multilocus variable number tandem repeat analysis primer; PCR primer; ss.  
XX OS Bacillus anthracis.  
XX PN WO200181543-A2.  
XX PD 01-NOV-2001.  
XX PF 26-APR-2001; 2001WO-US013373.  
XX PR 26-APR-2000; 2000US-0199911P.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Keim PS, Jackson PJ;  
XX DR WPI; 2002-055350/07.  
XX PT Novel nucleic acid molecules, primers for multilocus variable number  
PT tandem repeat analysis which are specific for Bacillus anthracis  
PT templates, useful for DNA finger printing of B. anthracis.  
XX PS Claim 43; Page 30; 71pp; English.  
XX CC The invention relates to an isolated nucleic acid molecule (ABA90786-  
CC ABA90790 and ABA90848-ABA90875) and a MLVA primer (multilocus VNTR  
CC (variable number tandem repeat) analysis primer, ABA90770-ABA90785 and  
CC ABA90791-ABA90847) specific for Bacillus anthracis templates used in DNA  
CC finger printing of B. anthracis, useful for identifying strains of B.  
CC anthracis and related bacteria  
XX SQ Sequence 2014 BP; 884 A; 278 C; 524 G; 328 T; 0 U; 0 Other;  
Query Match 82.0%; Score 16.4; DB 6; Length 2014;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAGTAAATGCAGAAACAG 19  
DB 419 AAGTAAATGCGAAACAG 436  
RESULT 58  
ABA90848  
ID ABA90848 standard; DNA; 2019 BP.

XX ABA90848;  
AC 14-FEB-2002 (first entry)  
XX  
XX Bacillus anthracis variable region DNA sequence SEQ ID NO 79.  
XX  
XX Bacillus anthracis; MLVA primer; VNTR; DNA finger printing; anthrax;  
KW multilocus variable number tandem repeat analysis primer; ds.  
XX  
XX Bacillus anthracis.  
OS  
XX WO200181543-A2.  
XX  
XX 01-NOV-2001.  
XX  
XX 26-APR-2001; 2001WO-US013373.  
XX  
XX 26-APR-2000; 2000US-0199911P.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Keim PS, Jackson PJ;  
XX  
XX WPI; 2002-055350/07.  
XX  
XX Novel nucleic acid molecules, primers for multilocus variable number  
PT tandem repeat analysis which are specific for Bacillus anthracis  
PT templates, useful for DNA finger printing of B. anthracis.  
XX  
XX Claim 47; Page 33-34; 71pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (ABA90786-  
CC ABA90790 and ABA90848-ABA90875) and a MLVA primer (multilocus VNTR  
CC (variable number tandem repeat) analysis primer, ABA90770-ABA90785 and  
CC ABA90791-ABA90847) specific for Bacillus anthracis templates used in DNA  
CC finger printing of B. anthracis, useful for identifying strains of B.  
CC anthracis and related bacteria  
CC  
XX Sequence 2019 BP; 884 A; 279 C; 524 G; 328 T; 0 U; 4 Other;  
SQ

Query Match 82.0%; Score 16.4; DB 6; Length 2019;  
Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 419 AAGTAAATGCAGAAACAG 436

RESULT 59  
ABQ67188/c  
ID ABQ67188 standard; DNA; 3807 BP.  
XX  
XX AC ABQ67188;  
XX  
XX 29-AUG-2002 (first entry)  
DT  
XX Listeria innocua contig DNA sequence #1.  
DE  
XX Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.  
XX  
XX Listeria innocua.  
OS  
XX WO200228891-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 04-OCT-2001; 2001WO-FR003061.  
XX  
XX 04-OCT-2000; 2000FR-00012697.  
XX

PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Kunet F, Glaser P;  
XX  
XX WPI; 2002-332479/37.  
XX  
XX New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators.  
XX  
XX Claim 5; SEQ ID NO 1; 180pp; French.  
XX  
XX The present invention relates to nucleic acid sequences (ABQ67188-  
CC ABQ71212) from Listeria sp. The sequences are useful as probes and  
CC primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of gene  
CC expression. Proteins encoded by the nucleic acid sequences can be used to  
CC screen for compounds that modulate gene expression, replication and  
CC pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in anti-  
CC Listeria vaccines. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3807 BP; 1392 A; 600 C; 641 G; 1174 T; 0 U; 0 Other;  
SQ

Query Match 82.0%; Score 16.4; DB 6; Length 3807;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 3464 AAGTAAATGCAGAAACAG 3447

RESULT 60  
ACA22058  
ID ACA22058 standard; DNA; 3912 BP.  
XX  
XX ACA22058;  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX Prokaryotic essential gene #3715.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Bacillus anthracis.  
OS  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR  
XX 06-SEP-2001; 2001US-00948993.  
PR  
XX 25-OCT-2001; 2001US-0342923P.  
PR  
XX 08-FEB-2002; 2002US-00072851.  
PR  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX P-PSDB; ABU18188.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 9928; 1766pp; English.

XX  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3912 BP; 1499 A; 605 C; 970 G; 838 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 8; Length 3912;

Best Local Similarity 94.4%; Pred. No. 1.5e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19

Db 1172 AAGTAAATGCAGAAACAG 1189

Search completed: January 27, 2006, 22:15:38  
Job time : 194.966 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 364.315 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattggtggttatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA Main:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	7	US-10-465-205-2
2	21	100.0	21	9	US-10-716-005-2
3	21	100.0	1323	7	US-10-465-205-12
4	21	100.0	1378	7	US-10-465-205-14
5	21	100.0	1379	7	US-10-465-205-7
6	21	100.0	1379	7	US-10-465-205-13
7	21	100.0	1379	7	US-10-465-205-16
8	21	100.0	1384	7	US-10-465-205-8
9	21	100.0	1384	7	US-10-465-205-10
10	21	100.0	1390	7	US-10-465-205-11
11	21	100.0	1393	7	US-10-465-205-15
12	21	100.0	1447	7	US-10-465-205-17
13	21	100.0	1734	7	US-10-282-122A-38442
14	21	100.0	1803	7	US-10-465-205-5
15	21	100.0	402	8	US-10-472-928-2351
16	19.4	92.4	1287	9	US-10-617-320-1469
17	19.4	92.4	1734	3	US-09-815-242-9183
18	19.4	92.4	1734	7	US-10-282-122A-37804
19	19.4	92.4	8195	2	US-08-961-527-94
20	19.4	92.4	8195	7	US-10-158-844-94
21	19.4	92.4	2162598	8	US-10-472-928-4979
22	19.4	92.4	553	6	US-10-029-386-1376
23	18.4	87.6			

C 24	18.4	87.6	1084	5	US-10-027-632-10931	Sequence 10931, A
C 25	18.4	87.6	1084	5	US-10-027-632-10932	Sequence 10932, A
C 26	18.4	87.6	1084	5	US-10-027-632-10933	Sequence 10933, A
C 27	18.4	87.6	1084	6	US-10-027-632-10931	Sequence 10931, A
C 28	18.4	87.6	1084	6	US-10-027-632-10932	Sequence 10932, A
C 29	18.4	87.6	1084	6	US-10-027-632-10933	Sequence 10933, A
C 30	17.8	84.8	1830	9	US-10-501-282-5273	Sequence 5273, Ap
C 31	17.8	84.8	1830	9	US-10-501-282-5275	Sequence 5275, Ap
C 32	17.8	84.8	1830	9	US-10-501-282-5277	Sequence 5277, Ap
C 33	17.8	84.8	1830	9	US-10-501-282-5279	Sequence 5279, Ap
C 34	17.8	84.8	2914	6	US-10-094-749-914	Sequence 914, Appl
C 35	17.8	84.8	2914	7	US-10-467-506A-15	Sequence 15, Appl
C 36	17.8	84.8	1754382	9	US-10-501-282-6651	Sequence 6651, Ap
C 37	17.4	82.9	25	8	US-10-719-900-73644	Sequence 73644, A
C 38	16.8	80.0	412	3	US-09-974-300-1665	Sequence 1665, Ap
C 39	16.8	80.0	532	4	US-09-925-065A-30849	Sequence 30849, A
C 40	16.8	80.0	532	4	US-09-925-065A-30850	Sequence 30850, A
C 41	16.8	80.0	616	4	US-09-925-065A-35263	Sequence 35263, A
C 42	16.8	80.0	619	4	US-09-925-065A-511481	Sequence 511481, A
C 43	16.8	80.0	675	4	US-09-925-065A-501042	Sequence 501042, A
C 44	16.8	80.0	1464	7	US-10-335-977-4387	Sequence 4387, Ap
C 45	16.8	80.0	1554	7	US-10-335-977-4388	Sequence 4388, Ap
C 46	16.8	80.0	2547	3	US-09-989-739-18	Sequence 18, Appl
C 47	16.8	80.0	10516	3	US-09-754-468-29	Sequence 29, Appl
C 48	16.8	80.0	51281	3	US-09-997-722-139	Sequence 139, Appl
C 49	16.8	80.0	160921	5	US-10-087-192-1672	Sequence 1672, Ap
C 50	16.4	78.1	520	4	US-09-925-065A-452149	Sequence 452149, A
C 51	16.4	78.1	531	4	US-09-925-065A-392178	Sequence 392178, A
C 52	16.4	78.1	643	4	US-09-925-065A-312506	Sequence 312506, A
C 53	16.4	78.1	647	4	US-09-925-065A-453189	Sequence 453189, A
C 54	16.4	78.1	647	4	US-09-925-065A-453190	Sequence 453190, A
C 55	16.4	78.1	111331	9	US-10-461-862-101	Sequence 101, Appl
C 56	16.2	77.1	164	6	US-10-029-386-24435	Sequence 24435, A
C 57	16.2	77.1	371	8	US-10-914-037-626	Sequence 626, Appl
C 58	16.2	77.1	434	3	US-09-783-590-9031	Sequence 9031, Ap
C 59	16.2	77.1	520	4	US-09-925-065A-358678	Sequence 358678, A
C 60	16.2	77.1	540	6	US-10-029-386-10735	Sequence 10735, A

ALIGNMENTS

RESULT 1  
US-10-465-205-2  
; Sequence 2, Application US/10465205  
; Publication NO. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhrl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: US/10/081,923  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-465-205-2

Query Match 100.0%; Score 21; DB 7; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21

Db 1 TGCATGTATGGGTATCTTCC 21

```

RESULT 2
US-10-716-005-2
; Sequence 2, Application US/10716005
; Publication No. US20050106578A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill III, Franklin R.
; APPLICANT: Aichinger, Christian
; APPLICANT: Reiser, Astrid
; TITLE OF INVENTION: Detection of Group B Streptococcus
; FILE REFERENCE: 07039/460001
; CURRENT APPLICATION NUMBER: US/10/716,005
; CURRENT FILING DATE: 2003-11-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-10-716-005-2

Query Match      100.0%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
      |||||
DB      1 TGCATGTATGGGTTATCTTCC 21

RESULT 3
US-10-465-205-12/c
; Sequence 12, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 2
US-10-465-205-12

Query Match      100.0%; Score 21; DB 7; Length 1323;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
      |||||
DB      193 TGCATGTATGGGTTATCTTCC 173

RESULT 4
US-10-465-205-14/c
; Sequence 14, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 6
US-10-465-205-14

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
      |||||
DB      209 TGCATGTATGGGTTATCTTCC 189

RESULT 5
US-10-465-205-7/c
; Sequence 7, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 6
US-10-465-205-7

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
      |||||
DB      209 TGCATGTATGGGTTATCTTCC 189

RESULT 6
US-10-465-205-13/c
; Sequence 13, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 6
US-10-465-205-13

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
      |||||
DB      209 TGCATGTATGGGTTATCTTCC 189

```

; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 4  
US-10-465-205-13

Query Match 100.0%; Score 21; DB 7; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189

RESULT 7  
US-10-465-205-16/c  
; Sequence 16, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 1379  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 1  
US-10-465-205-16

Query Match 100.0%; Score 21; DB 7; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 8  
US-10-465-205-8/c  
; Sequence 8, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1384  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 5  
US-10-465-205-8

Query Match 100.0%; Score 21; DB 7; Length 1384;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 9  
US-10-465-205-10/c  
; Sequence 10, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 1384  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 8  
US-10-465-205-10

Query Match 100.0%; Score 21; DB 7; Length 1384;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189

RESULT 10  
US-10-465-205-9/c  
; Sequence 9, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 1385  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 7  
US-10-465-205-9

Query Match 100.0%; Score 21; DB 7; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 11  
US-10-465-205-11/c  
; Sequence 11, Application US/10465205

```
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 9
US-10-465-205-11

Query Match      100.0%; Score 21; DB 7; Length 1390;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 224 TGCATGTATGGGTTATCTTCC 204

RESULT 12
US-10-465-205-15/c
; Sequence 15, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 3
US-10-465-205-15

Query Match      100.0%; Score 21; DB 7; Length 1393;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 13
US-10-465-205-17/c
; Sequence 17, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 11
US-10-465-205-17

Query Match      100.0%; Score 21; DB 7; Length 1447;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 14
US-10-282-122A-38442/c
; Sequence 38442, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38442
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-10-282-122A-38442

Query Match      100.0%; Score 21; DB 7; Length 1734;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 15

US-10-465-205-5/c  
; Sequence 5, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uh1, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-366001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: US/10/081,923  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptiI sequence from Oklahoma University M1 strain  
; PUBLICATION INFORMATION:  
; AUTHORS: Ferretti et al.  
; JOURNAL: Proc. Natl. Acad. Sci. USA  
; VOLUME: 98  
; PAGES: 4658-4663  
; DATE: 2001-01-01  
; US-10-465-205-5

Query Match 100.0%; Score 21; DB 7; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 16

US-10-472-928-2351/c  
; Sequence 2351, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH  
; FILE REFERENCE: P026926W0  
; CURRENT APPLICATION NUMBER: US/10/472,928  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: GB-0107658.7  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 2351  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; US-10-472-928-2351

Query Match 92.4%; Score 19.4; DB 8; Length 402;  
Best Local Similarity 95.2%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 17

US-10-617-320-1469/c  
; Sequence 1469, Application US/10617320  
; Publication No. US20050136404A1  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/617,320  
; FILING DATE: 10-Jul-2003  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1469:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1287 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...1287  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1469:  
US-10-617-320-1469

Query Match 92.4%; Score 19.4; DB 9; Length 1287;  
Best Local Similarity 95.2%; Pred. No. 21;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 431 TGCATGATGGTTATCTTCC 411

RESULT 18

US-09-815-242-9183/c  
; Sequence 9183, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 9183  
; LENGTH: 1734  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1734)  
US-09-815-242-9183

Query Match 92.4%; Score 19.4; DB 3; Length 1734;  
Best Local Similarity 95.2%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 377 TGCATGTATGGGTATCTTCC 357

RESULT 19  
US-10-282-122A-37804/c  
; Sequence 37804, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 37804  
; LENGTH: 1734  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-37804

Query Match 92.4%; Score 19.4; DB 7; Length 1734;  
Best Local Similarity 95.2%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 377 TGCATGTATGGGTATCTTCC 357

## RESULT 20

US-08-961-527-94/c  
; Sequence 94, Application US/08961527  
; Publication No. US2002003232A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunach  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8195 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-94

```
Query Match          92.4%; Score 19.4; DB 2; Length 8195;
Best Local Similarity 95.2%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 1384 TGCATGTATGGGTGTCCTCC 1364

RESULT 21
US-10-158-844-94/c
; Sequence 94, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-10-158-844-94

Query Match          92.4%; Score 19.4; DB 7; Length 8195;
Best Local Similarity 95.2%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 1384 TGCATGTATGGGTGTCCTCC 1364

RESULT 22
US-10-472-928-4979
; Sequence 4979, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
```

```
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4979
; LENGTH: 2162598
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-472-928-4979

Query Match          92.4%; Score 19.4; DB 8; Length 2162598;
Best Local Similarity 95.2%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 1008618 TGCATGTATGGGTGTCCTCC 1008638

RESULT 23
US-10-029-386-1376
; Sequence 1376, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1376
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21 70.0
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: SWISSPROT HIT: O14771, EVALUAE 4.00e-02
; OTHER INFORMATION: EST HUMAN HIT: AA228964.1, EVALUAE 2.00e-01
; OTHER INFORMATION: NT HIT: AF010322.1, EVALUAE 2.10e+00
US-10-029-386-1376

Query Match          87.6%; Score 18.4; DB 6; Length 553;
Best Local Similarity 95.0%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 20
Db 213 TGCATGTATGAGTTATCTTC 232

RESULT 24
US-10-027-632-10931/c
; Sequence 10931, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
```

```
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10931
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-10931
```

```
Query Match      87.6%; Score 18.4; DB 5; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGCATGTATGGGTATCTTC 20
   ||||| ||||| ||||| |||||
Db 693 TGCATGTATGAGTTATCTTC 674
```

```
RESULT 25
US-10-027-632-10932/c
; Sequence 10932, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10932
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-10932
```

```
Query Match      87.6%; Score 18.4; DB 5; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGCATGTATGGGTATCTTC 20
   ||||| ||||| ||||| |||||
Db 693 TGCATGTATGAGTTATCTTC 674
```

```
RESULT 26
US-10-027-632-10933/c
; Sequence 10933, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10933
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-10933
```

```
Query Match      87.6%; Score 18.4; DB 5; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGCATGTATGGGTATCTTC 20
   ||||| ||||| ||||| |||||
Db 693 TGCATGTATGAGTTATCTTC 674
```

```
RESULT 27
US-10-027-632-10931/c
; Sequence 10931, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10931
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-10931
```

```
Query Match      87.6%; Score 18.4; DB 6; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```





```
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5275
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (88)..(1827)
US-10-501-282-5275

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 32
US-10-501-282-5277/c
; Sequence 5277, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5277
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)..(1827)
US-10-501-282-5277

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 33
US-10-501-282-5279/c
; Sequence 5279, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
```

```
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5279
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1827)
US-10-501-282-5279

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 34
US-10-094-749-914/c
; Sequence 914, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAWATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 914
; LENGTH: 2914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-914

Query Match      84.8%; Score 17.8; DB 6; Length 2914;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 TGCATGTATGGGTTATCTTC 21  
|||  
pb 1517 TGCATGTATGGGTTTATTTC 1497

```

RESULT 35
US-10-467-506A-15/c
; Sequence 15, Application US/10467506A
; Publication No. US20040170994A1
; GENERAL INFORMATION:
; APPLICANT: Callen, David Fredrick
; APPLICANT: Kremmidiotis, Gabriel
; APPLICANT: Whitmore, Scott
; APPLICANT: Gardner, Alison
; APPLICANT: Powell, Jason
; TITLE OF INVENTION: DNA SEQUENCES FOR HUMAN TUMOUR SUPPRESSOR GENES
; FILE REFERENCE: 1386-15
; CURRENT APPLICATION NUMBER: US/10/467,506A
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/AU02/00138
; PRIOR FILING DATE: 2002-02-12
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 2914
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: cbs
; LOCATION: (333)..(1652)
US-10-467-506A-15

```

```
Query Match      84.8%; Score 17.8; DB 7; Length 2914;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 TGCATGATGGGTTATCTTC 21  
db 1517 TGCATGATGGGTTTATCTTC 1497

RESULT 36  
US-10-501-282-6651  
; Sequence 6651, Application US/10501282  
; Publication No. US20050203280A1  
; GENERAL INFORMATION:  
; APPLICANT: MCMICHAEL, JOHN CALHOUN  
; APPLICANT: ZAGORSKY, ROBERT JOHN  
; APPLICANT: RUSSELL, DAVID PARRISH  
; APPLICANT: FLETCHER, LEAH DIANE  
; TITLE OF INVENTION: ALLOIOCCOCUS OTITIDIS  
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF  
; FILE REFERENCE: AM100780 L2  
; CURRENT APPLICATION NUMBER: US/10/501,282  
; CURRENT FILING DATE: 2004-07-09  
; PRIOR APPLICATION NUMBER: 60/333,777  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 60/426,742  
; PRIOR FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: PCT/US02/36123  
; PRIOR FILING DATE: 2002-11-25  
; NUMBER OF SEQ ID NOS: 6653  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6651  
; LENGTH: 1754382  
; TYPE: DNA  
; ORGANISM: Alloiococcus otitidis  
US-10-501-282-6651

Query Match 84.8%; Score 17.8; DB 9; Length 1754382;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
|||  
Db 1394003 TGCATGTAGGGGTGTCTTCC 1394023

```

RESULT 37
US-10-719-900-73644
; Sequence 73644, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 73644
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-73644

```

Query Match 82.9%; Score 17.4; DB 8; Length 25;  
Best Local Similarity 94.7%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTTATCTTC 20  
|||||  
5 GCATGTATGGGTTATGTTTC 23

```

RESULT 38
US-09-974-300-1665
; Sequence 1665, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/690,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1665
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1665

```

Query Match	80.0%;	Score 16.8;	DB 3;	Length 412;
Best Local Similarity	90.0%;	Pred. No. 3.6e+02;		
Matches	18. Conservative	0. Mismatches	2. Indels	0. Gaps

QY 2 GCATGTATGGGTATCTCC 21  
|||  
78 GCATGTATGGGTATCTGCC 97  
|||

RESULT 39  
US-09-925-065A-30849/C  
; Sequence 30849, Application US/0925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30849
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-30849

Query Match      80.0%; Score 16.8; DB 4; Length 532;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      468 TGCATGTATGGATTCTTC 449

RESULT 40
US-09-925-065A-30850/c
; Sequence 30850, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30850
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-30850

Query Match      80.0%; Score 16.8; DB 4; Length 532;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      468 TGCATGTATGGATTCTTC 449

RESULT 41
US-09-925-065A-355263/c
; Sequence 355263, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 355263
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-355263

Query Match      80.0%; Score 16.8; DB 4; Length 616;
Best Local Similarity 90.0%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      206 TGCATGTATGGGTTATCATC 187

RESULT 42
US-09-925-065A-511481
; Sequence 511481, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 511481
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-511481

Query Match      80.0%; Score 16.8; DB 4; Length 619;
Best Local Similarity 90.0%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      469 TGCATGTATGGGTTATCATC 488
```



```
;
;
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1...1554
;   SEQUENCE DESCRIPTION: SEQ ID NO: 4388;
US-10-335-977-4388

Query Match      80.0%; Score 16.8; DB 7; Length 1554;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
    ||||| ||||| ||||| |||||
Db 1463 TGCATGTATGGGTATCTTC 1444

RESULT 46
US-09-989-739-18/c
; Sequence 18, Application US/09989739
; Publication No. US20030140364A1
; GENERAL INFORMATION:
; APPLICANT: HINCHEY, BRENDAN
; TITLE OF INVENTION: SONG, HEB-SOOK
; FILE REFERENCE: DEKM.177US
; CURRENT APPLICATION NUMBER: US/09/989,739
; CURRENT FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 2547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-989-739-18

Query Match      80.0%; Score 16.8; DB 3; Length 2547;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
    ||||| ||||| ||||| |||||
Db 1824 TGCATGTCTGGGTATCTTC 1805

RESULT 47
US-09-754-468-29/c
; Sequence 29, Application US/09754468
; Publication No. US20050192237A1
; GENERAL INFORMATION:
; APPLICANT: Iversen, Patrick L.
; TITLE OF INVENTION: Antisense Antibacterial Cell Division
; FILE REFERENCE: 0450-0033.30
; CURRENT APPLICATION NUMBER: US/09/754,468
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,484
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 10516
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-09-754-468-29

Query Match      80.0%; Score 16.8; DB 3; Length 10516;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
```

```
Db 5876 TGCATGTATGGGTATCTTC 5857
    ||||| ||||| ||||| |||||

RESULT 48
US-09-997-722-139/c
; Sequence 139, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 139
; LENGTH: 51281
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3889)..(5506)
; OTHER INFORMATION: "n" at positions 3889 through 5506 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12248)..(15107)
; OTHER INFORMATION: "n" at positions 12248 through 15107 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34241)..(34260)
; OTHER INFORMATION: "n" at positions 34241 through 34260 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (35928)..(35947)
; OTHER INFORMATION: "n" at positions 35928 through 35947 can be any base.
US-09-997-722-139

Query Match      80.0%; Score 16.8; DB 3; Length 51281;
Best Local Similarity 90.0%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21
    ||||| ||||| ||||| |||||
Db 37474 GCATGTGGGTATCTTC 37455

RESULT 49
US-10-087-192-1672
; Sequence 1672, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1672
; LENGTH: 160921
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-087-192-1672

Query Match      80.0%; Score 16.8; DB 5; Length 160921;
Best Local Similarity 90.0%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      138653 TGCATGTATGTGTTTCTTC 138672

RESULT 50
US-09-925-065A-452149/c
; Sequence 452149, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 452149
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-452149

Query Match      78.1%; Score 16.4; DB 4; Length 520;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      434 TCCATGTATGGGTTATCT 417

RESULT 51
US-09-925-065A-392178/c
; Sequence 392178, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 392178
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-392178

Query Match      78.1%; Score 16.4; DB 4; Length 531;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      435 TCCATGTATGGGTTATCT 418

RESULT 52
US-09-925-065A-312506/c
; Sequence 312506, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 312506
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-312506

Query Match      78.1%; Score 16.4; DB 4; Length 643;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      446 TGCATGTCTGGGTTATCT 429

RESULT 53
US-09-925-065A-453189/c
; Sequence 453189, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

```
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 453189
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-453189
```

```
Query Match          78.1%; Score 16.4; DB 4; Length 647;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 TGCATGTATGGGTTATCT 18
          ||||| ||||| |||||
Db      446 TGCATGTCTGGGTTATCT 429
```

```
RESULT 54
US-09-925-065A-453190/c
; Sequence 453190, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 453190
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-453190
```

```
Query Match          78.1%; Score 16.4; DB 4; Length 647;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 TGCATGTATGGGTTATCT 18
          ||||| ||||| |||||
Db      446 TGCATGTCTGGGTTATCT 429
```

```
RESULT 55
US-10-461-862-101/c
; Sequence 101, Application US/10461862
; Publication No. US20050090434A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 52945201800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 111331
; TYPE: DNA
; ORGANISM: Mus musculus
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(111331)
; OTHER INFORMATION: n = A,T,C or G
US-10-461-862-101
```

```
Query Match          78.1%; Score 16.4; DB 9; Length 111331;
Best Local Similarity 94.4%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 GCATGTATGGGTTATCTT 19
          ||||| ||||| |||||
Db      21925 GCATGTATGGGTTATCTT 21908
```

```
RESULT 56
US-10-029-386-24435
; Sequence 24435, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24435
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: NT HIT: AFI44387.1, EVALUAE 3.60e-02
; OTHER INFORMATION: SWISSPROT HIT: O28578, EVALUAE 4.40e+00
US-10-029-386-24435
```

```
Query Match          77.1%; Score 16.2; DB 6; Length 164;
Best Local Similarity 85.7%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 TGCATGTATGGGTTATCTTCC 21
          ||||| ||||| |||||
Db      104 TTCTGTATGGATTATCTTCC 124
```

```
RESULT 57
US-10-914-037-626/c
; Sequence 626, Application US/10914037
; Publication No. US20050003444A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NOVEL HUMAN POLYNUCLEOTIDES AND THE
; FILE REFERENCE: 8535-0029-999
; CURRENT APPLICATION NUMBER: US/10/914,037
; CURRENT FILING DATE: 2004-08-06
; PRIOR APPLICATION NUMBER: US/09/428,674
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: US 60/106,442
; PRIOR FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 626
; LENGTH: 371
```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-914-037-626

Query Match      77.1%; Score 16.2; DB 8; Length 371;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTATCTTCC 21
Db 70 TGCATGTATCTGTCTTCTCC 50

RESULT 58
US-09-783-590-9031/c
; Sequence 9031, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9031
; LENGTH: 494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (200)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (216)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (254)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (266)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (286)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (326)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (332)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (370)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (372)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (377)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (395)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (429)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (436)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (445)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (446)
; OTHER INFORMATION: n equals a,t,g, or c
```

; NAME/KEY: misc feature  
; LOCATION: (458)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (469)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (475)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (477)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (478)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (483)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (484)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-783-590-9031

Query Match 77.1%; Score 16.2; DB 3; Length 494;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 159 TGGATATATGGTTTCTTCC 139

RESULT 59  
US-09-925-065A-358678/c  
; Sequence 358678, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 358678  
; LENGTH: 520  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-358678

Query Match 77.1%; Score 16.2; DB 4; Length 520;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 25 TGCATGTATGGCTTCTTCC 5

RESULT 60  
US-10-029-386-10735  
; Sequence 10735, Application US/10029386

; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10735  
; LENGTH: 540  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR16.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
; OTHER INFORMATION: SWISSPROT HIT: P09065, EVALUE 3.90e+00  
; OTHER INFORMATION: NT HIT: A8004835.1, EVALUE 1.30e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AA131046.1, EVALUE 3.00e+00  
US-10-029-386-10735

Query Match 77.1%; Score 16.2; DB 6; Length 540;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 461 TTCTGTATGGATTATCTTCC 481

Search completed: January 28, 2006, 01:45:35  
Job time : 371.315 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 211.652 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattatgggttatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA New:  
1: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq3:  
11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.2	77.1	25	8	US-11-121-849-323547 Sequence 323547,
2	16.2	77.1	811	7	US-10-750-185-33099 Sequence 33099, A
3	16.2	77.1	811	7	US-10-750-623-33099 Sequence 33099, A
4	16.2	77.1	1712	7	US-10-750-185-57490 Sequence 57490, A
5	16.2	77.1	1712	7	US-10-750-623-57490 Sequence 57490, A
6	16.2	77.1	4975	7	US-10-750-185-37365 Sequence 37365, A
7	16.2	77.1	4975	7	US-10-750-623-37365 Sequence 37365, A
8	15.8	75.2	201	7	US-10-995-561-20395 Sequence 20395, A
9	15.8	75.2	600	7	US-10-750-185-2992 Sequence 2992, Ap
10	15.8	75.2	600	7	US-10-750-623-2992 Sequence 2992, Ap
11	15.8	75.2	1721	8	US-11-010-239-33 Sequence 33, Appl
12	15.8	75.2	380749	7	US-10-995-561-13216 Sequence 13216, A
13	15.8	75.2	1125000	7	US-10-995-561-13286 Sequence 13286, A
14	15.2	72.4	94	7	US-10-310-914A-19647 Sequence 19647, A
15	15.2	72.4	1094	7	US-10-750-185-55160 Sequence 55160, A
16	15.2	72.4	1094	7	US-10-750-623-55160 Sequence 55160, A
17	15.2	72.4	1406	7	US-10-750-185-55478 Sequence 55478, A
18	15.2	72.4	1406	7	US-10-750-623-55478 Sequence 55478, A
19	15.2	72.4	1789	7	US-10-750-185-45422 Sequence 45422, A
20	15.2	72.4	1789	7	US-10-750-623-45422 Sequence 45422, A
21	15.2	72.4	2234	7	US-10-750-185-56274 Sequence 56274, A
22	15.2	72.4	2234	7	US-10-750-623-56274 Sequence 56274, A

RESULT 1  
US-11-121-849-323547  
; Sequence 323547, Application US/11121849  
; Publication No. US20050272080A1  
; GENERAL INFORMATION:  
; APPLICANT: John Palma  
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded  
; TITLE OF INVENTION: Microarrays  
; FILE REFERENCE: 3684.1  
; CURRENT APPLICATION NUMBER: US/11/121.849  
; CURRENT FILING DATE: 2005-05-03  
; PRIOR APPLICATION NUMBER: 60/567,949  
; PRIOR FILING DATE: 2004-05-03  
; NUMBER OF SEQ ID NOS: 673904  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 323547  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-11-121-849-323547

Query Match 77.1%; Score 16.2; DB 8; Length 25;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21

Db 5 TGAATGTATGGTTCTTTTC 25

Sequence 59755, A  
Sequence 59755, A  
Sequence 203, App  
Sequence 50, Appl  
Sequence 89198, A  
Sequence 89198, A  
Sequence 200933, A  
Sequence 273497, A  
Sequence 25366, A  
Sequence 42892, A  
Sequence 42892, A  
Sequence 62244, A  
Sequence 62244, A  
Sequence 43249, A  
Sequence 43249, A  
Sequence 64265, A  
Sequence 64265, A  
Sequence 29709, A  
Sequence 29709, A  
Sequence 2197, Ap  
Sequence 13354, A  
Sequence 49, Appl  
Sequence 13233, A  
Sequence 6595, Ap  
Sequence 8804, Ap  
Sequence 1891, Ap  
Sequence 1139, Ap  
Sequence 2741, Ap  
Sequence 62709, A  
Sequence 62709, A  
Sequence 62508, A  
Sequence 62508, A  
Sequence 29432, A  
Sequence 29432, A  
Sequence 62216, A  
Sequence 62216, A  
Sequence 5, Appl

ALIGNMENTS

```
RESULT 2
US-10-750-185-33099
; Sequence 33099, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33099
; LENGTH: 811
; TYPE: DNA
; ORGANISM: Bovine 19866880817752
US-10-750-185-33099

Query Match      77.1%; Score 16.2; DB 7; Length 811;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21
||||| ||||||| |||||
Db 632 TGCATGGATGGGTTTCTACC 652

RESULT 3
US-10-750-623-33099
; Sequence 33099, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33099
; LENGTH: 811
; TYPE: DNA
; ORGANISM: Bovine 19866880817752
US-10-750-623-33099

Query Match      77.1%; Score 16.2; DB 7; Length 811;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21
||||| ||||||| |||||
Db 632 TGCATGGATGGGTTTCTACC 652

RESULT 4
US-10-750-185-57490
; Sequence 57490, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57490
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Bovine 19866880874779
US-10-750-185-57490

Query Match      77.1%; Score 16.2; DB 7; Length 1712;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21
||||| ||||||| |||||
Db 1314 TGCATATGTGGGTGATCTTCC 1334

RESULT 5
US-10-750-623-57490
; Sequence 57490, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57490
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Bovine 19866880874779
US-10-750-623-57490

Query Match      77.1%; Score 16.2; DB 7; Length 1712;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21
||||| ||||||| |||||
Db 1314 TGCATATGTGGGTGATCTTCC 1334

RESULT 6
US-10-750-185-37365/c
; Sequence 37365, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57490
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Bovine 19866880874779
US-10-750-185-37365/c

Query Match      77.1%; Score 16.2; DB 7; Length 1712;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21
||||| ||||||| |||||
Db 1314 TGCATATGTGGGTGATCTTCC 1334
```

```

; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37365
; LENGTH: 4975
; TYPE: DNA
; ORGANISM: Bovine 19866879992949
; US-10-750-185-37365

Query Match 77.1%; Score 16.2; DB 7; Length 4975;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGATGGTTATCTTCC 21
||||| ||||| |||||
DB 1760 TGCATGATGGGATATCTTTC 1740

RESULT 7
US-10-750-623-37365/c
; Sequence 37365, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37365
; LENGTH: 4975
; TYPE: DNA
; ORGANISM: Bovine 19866879992949
; US-10-750-623-37365

Query Match 77.1%; Score 16.2; DB 7; Length 4975;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGATGGTTATCTTCC 21
||||| ||||| |||||
DB 1760 TGCATGATGGGATATCTTTC 1740

RESULT 8
US-10-995-561-20395/c
; Sequence 20395, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37365
; LENGTH: 4975
; TYPE: DNA
; ORGANISM: Bovine 19866879992949
; US-10-750-623-37365

Query Match 77.1%; Score 16.2; DB 7; Length 4975;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGATGGTTATCTTCC 21
||||| ||||| |||||
DB 1760 TGCATGATGGGATATCTTTC 1740

RESULT 9
US-10-750-185-2992/c
; Sequence 2992, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2992
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT13234
; US-10-750-185-2992

Query Match 75.2%; Score 15.8; DB 7; Length 600;
Best Local Similarity 89.5%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGTTATCTT 19
||||| ||||| |||||
DB 425 TGTATATATGGTTATCTT 407

RESULT 10
US-10-750-623-2992/c
; Sequence 2992, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2992
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT13234
; US-10-750-623-2992

Query Match 75.2%; Score 15.8; DB 7; Length 600;
Best Local Similarity 89.5%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGTTATCTT 19
||||| ||||| |||||
DB 425 TGTATATATGGTTATCTT 407

```



Best Local Similarity 40.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
Db 64 UGGAUUUUUGGUUAUCUUC 83

RESULT 15  
US-10-750-185-55160/c  
; Sequence 55160, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55160  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Bovine 19866881169381  
US-10-750-185-55160

Query Match : 72.4%; Score 15.2; DB 7; Length 1094;  
Best Local Similarity 85.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
Db 611 GCAAGTATAGGTTTCTTCC 592

RESULT 16  
US-10-750-623-55160/c  
; Sequence 55160, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55160  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Bovine 19866881169381  
US-10-750-623-55160

Query Match : 72.4%; Score 15.2; DB 7; Length 1094;  
Best Local Similarity 85.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
Db 611 GCAAGTATAGGTTTCTTCC 592

RESULT 17  
US-10-750-185-55478  
; Sequence 55478, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55478  
; LENGTH: 1406  
; TYPE: DNA  
; ORGANISM: Bovine 19866880671997  
US-10-750-185-55478

Query Match : 72.4%; Score 15.2; DB 7; Length 1406;  
Best Local Similarity 85.0%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
Db 1082 TGAATGTATGTGTATCATC 1101

RESULT 18  
US-10-750-623-55478  
; Sequence 55478, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55478  
; LENGTH: 1406  
; TYPE: DNA  
; ORGANISM: Bovine 19866880671997  
US-10-750-623-55478

Query Match : 72.4%; Score 15.2; DB 7; Length 1406;  
Best Local Similarity 85.0%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
Db 1082 TGAATGTATGTGTATCATC 1101

RESULT 19  
US-10-750-185-45422  
; Sequence 45422, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 45422  
; LENGTH: 1789  
; TYPE: DNA  
; ORGANISM: Bovine 19866880353593  
US-10-750-185-45422

Query Match 72.4%; Score 15.2; DB 7; Length 1789;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 619 GAATGTATTGGTTATTTCC 638

RESULT 20  
US-10-750-623-45422  
; Sequence 45422, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 45422  
; LENGTH: 1789  
; TYPE: DNA  
; ORGANISM: Bovine 19866880353593  
US-10-750-623-45422

Query Match 72.4%; Score 15.2; DB 7; Length 1789;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 619 GAATGTATTGGTTATTTCC 638

RESULT 21

US-10-750-185-56274  
; Sequence 56274, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56274  
; LENGTH: 2234  
; TYPE: DNA  
; ORGANISM: Bovine 19866880566031  
US-10-750-185-56274

Query Match 72.4%; Score 15.2; DB 7; Length 2234;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 816 TGCATATATGGTTCTATTC 835

RESULT 22  
US-10-750-623-56274  
; Sequence 56274, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56274  
; LENGTH: 2234  
; TYPE: DNA  
; ORGANISM: Bovine 19866880566031  
US-10-750-623-56274

Query Match 72.4%; Score 15.2; DB 7; Length 2234;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTC 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 816 TGCATATATGGTTCTATTC 835

RESULT 23  
US-10-750-185-59755/c  
; Sequence 59755, Application US/10750185  
; Publication No. US20050260603A1



```
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 59755
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Bovine 19866880423655
US-10-750-185-59755

Query Match          72.4%; Score 15.2; DB 7; Length 2382;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTCC 21
Db      1168 GCATTAATGGTGTCTTCC 1149

RESULT 24
US-10-750-623-59755/c
; Sequence 59755, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 59755
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Bovine 19866880423655
US-10-750-623-59755

Query Match          72.4%; Score 15.2; DB 7; Length 2382;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTCC 21
Db      1168 GCATTAATGGTGTCTTCC 1149

RESULT 25
US-11-117-187-203
; Sequence 203, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 203
; LENGTH: 96583
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-203

Query Match          72.4%; Score 15.2; DB 8; Length 96583;
Best Local Similarity 85.0%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGCATGTATGGTTATCTTCC 20
Db      38324 TGCATGTATGGTTTCTTCC 38343

RESULT 26
US-11-121-086-50/c
; Sequence 50, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 50
; LENGTH: 203467
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-50

Query Match          72.4%; Score 15.2; DB 8; Length 203467;
Best Local Similarity 85.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTCC 21
Db      90768 GCATGCTTAGTTATCTTCC 90749

RESULT 27
US-11-101-244-89198/c
; Sequence 89198, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
```

```
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 89198
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-89198
```

```
Query Match      70.5%; Score 14.8; DB 9; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      18 ATGCTGTGTTATCTTCC 1
```

```
RESULT 28
US-11-083-784-89198/c
; Sequence 89198, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 89198
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-89198
```

```
Query Match      70.5%; Score 14.8; DB 10; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      18 ATGCTGTGTTATCTTCC 1
```

```
RESULT 29
US-11-121-849-200933/c
; Sequence 200933, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 200933
; LENGTH: 25
```

```
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-200933
```

```
Query Match      70.5%; Score 14.8; DB 8; Length 25;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 TGCATGTATGGGTTATCT 18
      ||||| ||||| |||||
Db      24 TGCATGTATGGGTTCT 7
```

```
RESULT 30
US-11-121-849-273497
; Sequence 273497, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 273497
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-273497
```

```
Query Match      70.5%; Score 14.8; DB 8; Length 25;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      6 ATGCTGTGATTATCTTCC 23
```

```
RESULT 31
US-10-995-561-25366/c
; Sequence 25366, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25366
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-25366
```

```
Query Match      70.5%; Score 14.8; DB 7; Length 201;
Best Local Similarity 88.9%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      121 ATGCATGGGTTAGCTTCC 104
```

```
RESULT 32
```

US-10-750-185-42892/c  
; Sequence 42892, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 42892  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Bovine 1986680852756  
US-10-750-185-42892

Query Match 70.5%; Score 14.8; DB 7; Length 774;  
Best Local Similarity 88.9%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTT 19  
||||| ||||| |||||  
Db 624 GCATGTGTGGGTTTCTT 607

RESULT 33  
US-10-750-623-42892/c  
; Sequence 42892, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 42892  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Bovine 1986680852756  
US-10-750-623-42892

Query Match 70.5%; Score 14.8; DB 7; Length 774;  
Best Local Similarity 88.9%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTT 19  
||||| ||||| |||||  
Db 624 GCATGTGTGGGTTTCTT 607

RESULT 34  
US-10-750-185-62244  
; Sequence 62244, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 62244  
; LENGTH: 982  
; TYPE: DNA  
; ORGANISM: Bovine 1986688112236  
US-10-750-185-62244

Query Match 70.5%; Score 14.8; DB 7; Length 982;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTCC 21  
||||| ||||| |||||  
Db 156 ATGATGTGATATCTTCC 173

RESULT 35  
US-10-750-623-62244  
; Sequence 62244, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 62244  
; LENGTH: 982  
; TYPE: DNA  
; ORGANISM: Bovine 1986688112236  
US-10-750-623-62244

Query Match 70.5%; Score 14.8; DB 7; Length 982;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTCC 21  
||||| ||||| |||||  
Db 156 ATGATGTGATATCTTCC 173

RESULT 36  
US-10-981-334-45  
; Sequence 45, Application US/10981334  
; Publication No. US2006008816A1  
; GENERAL INFORMATION:  
; APPLICANT: LU, Yu-Ping et al.  
; TITLE OF INVENTION: PROMOTER, PROMOTER CONTROL ELEMENTS, AND COMBINATIONS, AND USES

```
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 2750-1561PUS2
; CURRENT APPLICATION NUMBER: US/10/981,334
; CURRENT FILING DATE: 2004-11-04
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 45
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Promoter and/or promoter control element identified from
; OTHER INFORMATION: Arabidopsis thaliana or Oryza sativa.
US-10-981-334-45

Query Match          70.5%; Score 14.8; DB 6; Length 999;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTTATCTTC 20
   ||||| |||||
Db 848 CATTATGGATTATCTTC 865

RESULT 37
US-10-750-185-43249
; Sequence 43249, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43249
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Bovine 19866881139126
US-10-750-185-43249

Query Match          70.5%; Score 14.8; DB 7; Length 1338;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTTATCTTC 21
   ||||| |||||
Db 205 ATGCATGGGTTATTTCC 222

RESULT 38
US-10-750-623-43249
; Sequence 43249, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
```

```
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43249
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Bovine 19866881139126
US-10-750-623-43249

Query Match          70.5%; Score 14.8; DB 7; Length 1338;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTTATCTTC 21
   ||||| |||||
Db 205 ATGCATGGGTTATTTCC 222

RESULT 39
US-10-750-185-64265
; Sequence 64265, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64265
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Bovine 19866881415571
US-10-750-185-64265

Query Match          70.5%; Score 14.8; DB 7; Length 1560;
Best Local Similarity 88.9%; Pred. No. 2.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTTATCTTC 20
   ||||| |||||
Db 1129 CATTATGGGTTATCTAC 1146

RESULT 40
US-10-750-623-64265
; Sequence 64265, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
```

; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 64265  
; LENGTH: 1560  
; TYPE: DNA  
; ORGANISM: Bovine 19866881415571  
US-10-750-623-64265

Query Match 70.5%; Score 14.8; DB 7; Length 1560;  
Best Local Similarity 88.9%; Pred. No. 2.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTTATCTTC 20  
DB 1129 CATTATGGGTTATCTAC 1146

RESULT 41  
US-10-750-185-29709  
; Sequence 29709, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 29709  
; LENGTH: 1784  
; TYPE: DNA  
; ORGANISM: Bovine 19866880728339  
US-10-750-185-29709

Query Match 70.5%; Score 14.8; DB 7; Length 1784;  
Best Local Similarity 88.9%; Pred. No. 3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTTATCTTC 21  
DB 256 ATGTCTGGGTTCTCTCC 273

RESULT 42  
US-10-750-623-29709  
; Sequence 29709, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1

; SEQ ID NO 29709  
; LENGTH: 1784  
; TYPE: DNA  
; ORGANISM: Bovine 19866880728339  
US-10-750-623-29709

Query Match 70.5%; Score 14.8; DB 7; Length 1784;  
Best Local Similarity 88.9%; Pred. No. 3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTTATCTTC 21  
DB 256 ATGTCTGGGTTCTCTCC 273

RESULT 43  
US-11-136-527-2197  
; Sequence 2197, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIN version 3.2  
; SEQ ID NO 2197  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-2197

Query Match 70.5%; Score 14.8; DB 8; Length 4689;  
Best Local Similarity 88.9%; Pred. No. 3.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTTATCTTC 20  
DB 2768 CAGGTATGGGTTATCTTC 2785

RESULT 44  
US-10-995-561-13354/c  
; Sequence 13354, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13354  
; LENGTH: 47444  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(47444)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-995-561-13354

Query Match 70.5%; Score 14.8; DB 7; Length 47444;  
Best Local Similarity 88.9%; Pred. No. 5.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      4 ATGATGGGTTATCTTCC 21
Db      2089 ATTATGGTTATCTTCC 2072

RESULT 45
US-11-121-086-49/c
; Sequence 49, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 49
; LENGTH: 159146
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-49

Query Match      70.5%; Score 14.8; DB 8; Length 159146;
Best Local Similarity 88.9%; Pred. No. 6.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCT 18
Db      98113 TGAATGTATGGGTGTCT 98096

RESULT 46
US-10-995-561-13233/c
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      70.5%; Score 14.8; DB 7; Length 199130;
Best Local Similarity 88.9%; Pred. No. 6.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGATGGGTTATCTTCC 21
Db      190002 ATGCATGGTTAGTCTCC 189985

RESULT 47
US-11-121-849-6595/c
; Sequence 6595, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
```

```
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded ;
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6595
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-6595

Query Match      69.5%; Score 14.6; DB 8; Length 25;
Best Local Similarity 81.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
Db      21 TACATGCATGGGTTTCTGCC 1

RESULT 48
US-11-124-368A-8804/c
; Sequence 8804, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8804
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-8804

Query Match      69.5%; Score 14.6; DB 8; Length 201;
Best Local Similarity 81.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
Db      187 TGAATGCATGGGTTTGTCTCC 167

RESULT 49
US-10-793-626-1891/c
; Sequence 1891, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
```

Query Match	Score	DB	Length	Indels	Mismatches	Gaps
SEQ ID NO 1891	69.5%	14.6	7	0	0	0
LENGTH: 868						
TYPE: DNA						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: synthetic						
OTHER INFORMATION: nucleic acid sequence						
US-10-793-626-1891						
Query Match	69.5%	14.6	7	0	0	0
Best Local Similarity	81.0%	3.3e+02				
Matches	17	Conservative	0	Mismatches	4	Indels
Qy	1	TGCATGTATGGTTATCTTCC	21			
Db	272	TGCACGTCTGAATTAATCTTCC	252			
RESULT 50						
US-10-793-626-1139/c						
Sequence 1139, Application US/10793626						
Publication No. US20050255478A1						
GENERAL INFORMATION:						
APPLICANT: KIMMERLY, WILLIAM JOHN						
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS						
FILE REFERENCE: PU3480US						
CURRENT APPLICATION NUMBER: US/10/793,626						
CURRENT FILING DATE: 2004-03-04						
PRIOR APPLICATION NUMBER: 60/164,258						
PRIOR FILING DATE: 1999-11-09						
NUMBER OF SEQ ID NOS: 4472						
SOFTWARE: PatentIn Ver. 2.1						
SEQ ID NO 1139						
LENGTH: 912						
TYPE: DNA						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: synthetic						
OTHER INFORMATION: nucleic acid sequence						
US-10-793-626-1139						
Query Match	69.5%	14.6	7	0	0	0
Best Local Similarity	81.0%	3.4e+02				
Matches	17	Conservative	0	Mismatches	4	Indels
Qy	1	TGCATGTATGGTTATCTTCC	21			
Db	104	TGCACGTCTGAATTAATCTTCC	84			
RESULT 51						
US-10-793-626-2741/c						
Sequence 2741, Application US/10793626						
Publication No. US20050255478A1						
GENERAL INFORMATION:						
APPLICANT: KIMMERLY, WILLIAM JOHN						
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS						
FILE REFERENCE: PU3480US						
CURRENT APPLICATION NUMBER: US/10/793,626						
CURRENT FILING DATE: 2004-03-04						
PRIOR APPLICATION NUMBER: 60/164,258						
PRIOR FILING DATE: 1999-11-09						
NUMBER OF SEQ ID NOS: 4472						
SOFTWARE: PatentIn Ver. 2.1						
SEQ ID NO 2741						
LENGTH: 1080						
TYPE: DNA						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: synthetic						
OTHER INFORMATION: nucleic acid sequence						
US-10-793-626-2741						

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 41 TACATGTATGGGTCTCTCC 21

RESULT 54  
US-10-750-185-62508  
; Sequence 62508, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62508  
; LENGTH: 1208  
; TYPE: DNA  
; ORGANISM: Bovine 19866880913049  
US-10-750-185-62508

Query Match 69.5%; Score 14.6; DB 7; Length 1208;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 1037 TGCATTCATGGGCTTTCTTCC 1057

RESULT 55  
US-10-750-623-62508  
; Sequence 62508, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62508  
; LENGTH: 1208  
; TYPE: DNA  
; ORGANISM: Bovine 19866880913049  
US-10-750-623-62508

Query Match 69.5%; Score 14.6; DB 7; Length 1208;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 1037 TGCATTCATGGGCTTTCTTCC 1057

Db 1037 TGCATTCATGGGCTTTCTTCC 1057

RESULT 56  
US-10-750-185-29432  
; Sequence 29432, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29432  
; LENGTH: 1275  
; TYPE: DNA  
; ORGANISM: Bovine 19866881425942  
US-10-750-185-29432

Query Match 69.5%; Score 14.6; DB 7; Length 1275;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCATTCATGGGTATCTTCC 21  
Db 997 TGAATGTATGGGTTTTTTTC 1017

RESULT 57  
US-10-750-623-29432  
; Sequence 29432, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29432  
; LENGTH: 1275  
; TYPE: DNA  
; ORGANISM: Bovine 19866881425942  
US-10-750-623-29432

Query Match 69.5%; Score 14.6; DB 7; Length 1275;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCATTCATGGGTATCTTCC 21  
Db 997 TGAATGTATGGGTTTTTTTC 1017



RESULT 58

US-10-750-185-62216  
; Sequence 62216, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62216  
; LENGTH: 1353  
; TYPE: DNA  
; ORGANISM: Bovine 19866880904068  
US-10-750-185-62216

Query Match 69.5%; Score 14.6; DB 7; Length 1353;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 299 TCGTTGAATGCCCTTATCTTCC 319

RESULT 59

US-10-750-623-62216  
; Sequence 62216, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62216  
; LENGTH: 1353  
; TYPE: DNA  
; ORGANISM: Bovine 19866880904068  
US-10-750-623-62216

Query Match 69.5%; Score 14.6; DB 7; Length 1353;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 299 TCGTTGAATGCCCTTATCTTCC 319

RESULT 60

US-10-512-109-5/c  
; Sequence 5, Application US/10512109

; Publication No. US20050255546A1

; GENERAL INFORMATION:  
; APPLICANT: KIRIN BEER KABUSHIKI KAISHA  
; TITLE OF INVENTION: POLYPEPTIDE HAVING AN ACTIVITY TO SUPPORT PROLIFERATION OR SURVIVAL OF HEMATOPOIETIC PROGENITOR CELL, AN  
; TITLE OF INVENTION: OF HEMATOPOIETIC STEM CELL OR HEMATOPOIETIC PROGENITOR CELL, AN  
; TITLE OF INVENTION: FOR THE SAME  
; FILE REFERENCE: 905WO10P1572  
; CURRENT APPLICATION NUMBER: US/10/512,109  
; CURRENT FILING DATE: 2004-10-21  
; PRIOR APPLICATION NUMBER: US 60/376,001  
; PRIOR FILING DATE: 2002-04-26  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 1420  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-512-109-5

Query Match 69.5%; Score 14.6; DB 7; Length 1420;  
Best Local Similarity 81.0%; Pred. No. 3.7e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 1195 TTCATATATGGCTTATCTTGC 1175

Search completed: January 28, 2006, 02:00:40  
Job time : 215.902 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 2026.12 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-3  
Perfect score: 25  
Sequence: 1 caaattaaagagactattcgtaa 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues  
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gsa1:\*  
10: gb\_gsa2:\*  
11: gb\_gsa3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.8	79.2	281	10	CZ594841 CMHD-GT 2
C 2	19.8	79.2	388	10	CZ317624 ZMMBF0017
C 3	19.2	76.8	394	10	CG838808 ZMMBc025
C 4	19.2	76.8	399	10	CL453893 ZMMBb048
C 5	19.2	76.8	473	9	AZ932227 47A_dhz96
C 6	19.2	76.8	565	10	CZ687653 OA_Bba016
C 7	19.2	76.8	582	3	BM331796 MEST166-G
C 8	19.2	76.8	600	3	BM331796 MEST332-G
C 9	19.2	76.8	612	11	CR340097 mte1-71J2
C 10	19.2	76.8	618	3	BM349126 MEST308-G
C 11	19.2	76.8	642	11	DE026109 Branchios
C 12	19.2	76.8	673	3	BM348442 MEST230-H
C 13	19.2	76.8	765	10	CL680700 PRI012a.D
C 14	19.2	76.8	785	10	CNS02PKY Tetraodon
C 15	19.2	76.8	804	11	CR146965 Forward s
C 16	19.2	76.8	810	9	BZ784243 PUCAT94TB
C 17	19.2	76.8	812	11	CR173895 Forward s
C 18	19.2	76.8	815	10	CG348543 OGOFUL3TV
C 19	19.2	76.8	816	10	CG348533 OGOFUL3TH
C 20	19.2	76.8	872	9	CG681145 OGUFK42TV
C 21	19.2	76.8	895	2	BG479888 602527434
C 22	19.2	76.8	925	9	CG681138 OGUFK42TH

C 23	19.2	76.8	991	9	CC686212	CC686212
C 24	18.8	75.2	444	10	CG033706	CG033706
C 25	18.8	75.2	460	9	AQ989224	AQ989224
C 26	18.6	75.2	612	9	AQ233842	AQ233842
C 27	18.6	74.4	148	7	CV709364	CV709364
C 28	18.6	74.4	206	9	AZ794021	AZ794021
C 29	18.6	74.4	537	11	D8043516	D8043516
C 30	18.6	74.4	569	9	BZ615728	BZ615728
C 31	18.6	74.4	586	6	CB092313	CB092313
C 32	18.6	74.4	600	8	CX543634	CX543634
C 33	18.6	74.4	604	8	CX644973	CX644973
C 34	18.6	74.4	608	8	CX644972	CX644972
C 35	18.6	74.4	666	7	CJ340874	CJ340874
C 36	18.6	74.4	670	3	BI390310	BI390310
C 37	18.6	74.4	670	9	CC373687	CC373687
C 38	18.6	74.4	691	1	AW333363	AW333363
C 39	18.6	74.4	706	8	CX546303	CX546303
C 40	18.6	74.4	708	7	CJ397741	CJ397741
C 41	18.6	74.4	712	7	CV509866	CV509866
C 42	18.6	74.4	712	10	CZ287609	CZ287609
C 43	18.6	74.4	717	7	CV706617	CV706617
C 44	18.6	74.4	727	7	CV709207	CV709207
C 45	18.6	74.4	731	7	CJ344248	CJ344248
C 46	18.6	74.4	732	7	CJ357933	CJ357933
C 47	18.6	74.4	733	8	CX641978	CX641978
C 48	18.6	74.4	740	10	CZ279576	CZ279576
C 49	18.6	74.4	742	7	CJ361428	CJ361428
C 50	18.6	74.4	745	5	BUI30653	BUI30653
C 51	18.6	74.4	754	7	CJ358434	CJ358434
C 52	18.6	74.4	758	9	BZ246552	BZ246552
C 53	18.6	74.4	764	7	CV708436	CV708436
C 54	18.6	74.4	788	8	CX644101	CX644101
C 55	18.6	74.4	793	7	CV706539	CV706539
C 56	18.6	74.4	798	7	CV705522	CV705522
C 57	18.6	74.4	798	9	BZ536243	BZ536243
C 58	18.6	74.4	810	7	CV704979	CV704979
C 59	18.6	74.4	816	7	CV708661	CV708661
C 60	18.6	74.4	823	7	CJ401436	CJ401436

ALIGNMENTS

RESULT 1  
CZ594841/c  
LOCUS  
DEFINITION

CZ594841 281 bp mRNA linear GSS 30-JUN-2005  
CMHD-GT\_220D12-3 GTL\_R1\_pGTNMF Mus musculus cDNA clone  
CMHD-GT\_220D12-3 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CZ594841 GI:68367780  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 281)  
Stanford, W.L.  
www.cmhd.ca  
Unpublished (2002)  
Contact: Stanford WL  
Institute of Biomaterials & Biomedical Engineering  
University of Toronto  
407 Rosebrugh Bldg., 4 Taddle Creek Rd., Toronto, Ontario, Canada  
M5S 3G9  
Tel: 416 946 8379  
Fax: 416 978 4317  
Email: william.stanford@utoronto.ca

pGTNMF Gene trap insertion. The sequence tag is generated by 3' race. The ES cell line harboring this insertion of the target gene is available through the following web site:  
http://pokey.ihme.utoronto.ca/sequence\_report.php?id=220D12.  
Class: Gene Trap.



```

Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@wakeman.rutgers.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 104.
      Location/Qualifiers
          source            1. .399
              /organism="Zea mays"
              /mol_type="genomic DNA"
              /cultivar="B73"
              /db_xref="taxon:4577"
              /clone="ZMWBB0485M20"
              /lab_host="E. coli DH10B"
              /clone_lib="ZMWBBb (HindIII)"
              /note="Vector: pCUG1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match          76.8%;   Score 19.2;   DB 10;   Length 399;
Best Local Similarity 87.5%;   Pred. No. 5.8e+02;
Matches 21;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCA 24
    |||||
Db 212 CAAATTAAATCGCACTATCGTGCA 189

```

RESULT 5	A29332227	473 bp	DNA	linear	GSS 01-APR-2001
A29332227					
LOCUS	474.dhz96c07.s1		Saccharomyces unisporus	NRRL Y-1556	Saccharomyces
DEFINITION	unisporus genomic clone 474.dhz96c07.s1,		genomic survey sequence.		
ACCESSION	A29332227				
VERSION	A29332227.1	GI:13503139			
KEYWORDS	GSS.				
SOURCE	Saccharomyces unisporus				
ORGANISM	Saccharomyces unisporus				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 473)				
AUTHORS	Clifton, P., Hillier, L.W., Fulton, L., Graves, T., Miner, T.,				
	Gish, W.R., Waterston, R.H. and Johnston, M.				
TITLE	Surveying Saccharomyces genomes to identify functional elements by				
	comparative DNA sequence analysis				

```

FEATURES
  source
    1. 473
      Location/Qualifiers
        /organism="Saccharomyces unisporus"
        /mol_type="genomic DNA"
        /strain="NRRL Y-1556 (CBS 398)"
        /db_xref="taxon:27294"
        /clone="474.dhz96c07.s1"
        /clone_lib="Saccharomyces unisporus NRRL Y-1556"
        /note="Random genomic sequence"

ORIGIN

Query Match          76.8%; Score 19.2; DB 9; Length 473;
Best Local Similarity 87.5%; Pred. No. 5.8e+02;
Matches 21: Conservative 0; Mismatches 3; Indels 0; Gaps 0

```

<b>Qy</b>	<b>2</b>	<b>AAATTAAGAGACTATTTCGTGCAA</b>	<b>25</b>
<b>D<sub>b</sub></b>	<b>335</b>	<b>AAATTTAAAGAGATTATACATGCAA</b>	<b>358</b>

RESULT 6	
CZ687653/c	
LOCUS	565 bp DNA linear GSS 14-JUL-2005
DEFINITION	OA_BB0164M21.r OA_BBa Oryza alta genomic clone OA_BB0164M21 3', genomic survey sequence.
ACCESSION	CZ687653
VERSION	CZ687653.1
KEYWORDS	GI:70808805
SOURCE	GSS.
ORGANISM	Oryza alta
	Oryza alta
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1 (bases 1 to 565)
AUTHORS	SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wlasotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
TITLE	OMAP (Oryza Map Alignment Project) - Purdue University
COMMENT	Unpublished (2004) Contact: Scott A. Jackson Jackson Laboratory Purdue University 915 W. State St., West Lafayette, IN 47907, USA Tel: 7654963621 Fax: 7654967255

```

from the raw sequence read by clipping with Lucy version 1.198.
Bases 114-678 of the raw sequence (length 1100) were retained after
clipping.
Plate: 0164 row: M column: 21
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
1. .565
/organism="Oryza alta"
/mol_type="genomic DNA"
/db_xref="taxon:52545"
/clone="OA_BBa0164M21"
/tissue_type="young leaves"
/lab_host="DH10B-T1 phage resistant"
/clone_lib="OA_BBa"
/note="vector: pAG1BAC1; Site_1: HindIII; Site_2: HindIII"

```

RESULT	7
BM331796	
LOCUS	
DEFINITION	BM331796 linear EST 16-JAN-2000 MEST166-G03.T3 ISUM5-RN Zea mays cDNA clone MEST166-G03 3', mRNA sequence.
ACCESSION	BM331796
VERSION	BM331796.1 GI:18161957
KEYWORDS	EST.
SOURCE	Zea mays
ORGANISM	Zea mays Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

**AUTHORS** Wen, T. J., Qiu, F., Guo, L., Ashlock, D. A. and Schnable, P. S.  
**TITLE** Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Patrick S. Schnable

Schnable Laboratory  
Iowa State University  
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA  
Tel: 515-294-0975  
Fax: 515-294-5256  
Email: schnable@iastate.edu  
Individual basecall and confidence value were assigned using the Phred software,  
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>)#b  
rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.tigr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.  
PCR Primers  
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

#### FEATURES

source  
1..582  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST166-G03"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/clone\_lib="ISUM5-RN"  
/note="Vector: pT7T3PAC; Site 1: EcoRI; Site 2: NotI;  
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG), Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated first ear, ear shank, etiolated seedlings, callus, Cycloheximide-treated callus, Anaerobic treated seedlings, NAA (a-Naphthalene acetic acid)-treated seedlings, Kinetin-treated seedlings, ACPC (1-aminocyclopropane-1-carboxylic acid)-treated seedlings, Brassinolide-treated seedlings, ABA (Abscissic acid)-treated seedlings, GA (Gibberellic acid)-treated seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5', AACTGGAAGATTCGGCCGACGAGGAATTTTTTTTTTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT7T3PAC vector. The library then went through one round of normalization to Cot value of 5 based on the methods of Marcello Bento Soares (Genome Research 6: 791-806, 1996)."

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 582;  
Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 CAAATTAAGAGACTATTCGTGCA 24  
||||||| ||||| ||||| |||||  
Db 128 CAAATTAAGACTATGCGTGCA 151  
RESULT 8  
BM341275 600 bp mRNA linear EST 16-JAN-2002  
LOCUS MEST332-G07.T3 ISUM5-RN Zea mays cDNA clone MEST332-G07 3', mRNA  
DEFINITION sequence.

#### ACCESSION VERSION KEYWORDS SOURCE ORGANISM

BM341275  
BM341275.1 GI:18171435  
EST.  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 600)

#### REFERENCE

Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.  
Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones  
Unpublished (2001)

#### JOURNAL

#### COMMENT

Contact: Patrick S. Schnable  
Schnable Laboratory  
Iowa State University  
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA  
Tel: 515-294-0975  
Fax: 515-294-5256  
Email: schnable@iastate.edu  
Individual basecall and confidence value were assigned using the Phred software,  
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>)#b  
rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.tigr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.  
PCR Primers  
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

#### FEATURES

##### source

1..600  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST332-G07"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/clone\_lib="ISUM5-RN"  
/note="Vector: pT7T3PAC; Site 1: EcoRI; Site 2: NotI;  
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG), Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated first ear, ear shank, etiolated seedlings, callus, Cycloheximide-treated callus, Anaerobic treated seedlings, NAA (a-Naphthalene acetic acid)-treated seedlings, Kinetin-treated seedlings, ACPC (1-aminocyclopropane-1-carboxylic acid)-treated seedlings, Brassinolide-treated seedlings, ABA (Abscissic acid)-treated seedlings, GA (Gibberellic acid)-treated seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5', AACTGGAAGATTCGGCCGACGAGGAATTTTTTTTTTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT7T3PAC vector. The library then went through one round of normalization to Cot value of 5 based on the methods of Marcello Bento Soares (Genome Research 6: 791-806, 1996)."

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 600;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 127 CAAATTAAATGGACTATGCGTGCA 150

RESULT 9  
 CR340097/c  
 LOCUS  
 DEFINITION mtei-71J2RM1 BAC end, cultivar Jemalong A17 of Medicago truncatula,  
 genomic survey sequence.

ACCESSION CR340097  
 VERSION CR340097.1 GI:44910432  
 KEYWORDS GSS.

ORGANISM  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE  
 AUTHORS 1 (bases 1 to 612)  
 Genoscope.

TITLE Direct Submission  
 JOURNAL Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

FEATURES  
 source  
 1..612  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="Jemalong A17"  
 /db\_xref="taxon:3880"  
 /clone\_lib="MTEI"  
 /notes="Vector: pindigobac ; Site\_1: EcoRI ; Site\_2: EcoRI  
 ; Debelle F. and Chalhou B.  
 Genoscope sequence ID : mtei-71J2RM1"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 612;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 355 CAAATTAAATATCTATTCTGTGCA 332

RESULT 10  
 BM349126  
 LOCUS  
 DEFINITION BM3T308-G10.T3 ISUM5-RN Zea mays cDNA clone MEST308-G10 3', mRNA  
 sequence.

ACCESSION BM349126  
 VERSION BM349126.1 GI:18173738  
 KEYWORDS EST.

SOURCE  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 618)

Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.  
 Expressed Sequence Tags from B73 Maize: various stages and tissues  
 including seedlings treated with a variety of hormones  
 Unpublished (2001)

JOURNAL  
 COMMENT Contact: Patrick S. Schnable  
 Schnable Laboratory  
 Iowa State University  
 2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA  
 Tel: 515-294-0975

Fax: 515-294-5256  
 Email: schnable@iastate.edu  
 Individual basecall and confidence value were assigned using the  
 Phred software.

(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b  
 rt). Overall sequence quality assessment and vector trimming were  
 conducted using the Lucy software (<http://www.tigr.org/softlab/>).  
 Lucy parameters were set to ensure an overall trimmed quality of  
 97.5% or better without any vector fragments in the chosen  
 high-quality region of each sequence. Low-quality bases between the  
 poly-T and the high-quality region were replaced with N's to serve  
 as spacers.

PCR Primers

FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)

BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

## FEATURES

Location/Qualifiers  
 1..618  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="MEST308-G10"  
 /tissue\_type="mixed"  
 /lab\_host="DH10B"  
 /clone\_lib="ISUM5-RN"  
 /note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
 DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear  
 (0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk,  
 unpollinated first ear, ear shank, etiolated seedlings,  
 callus, Cycloheximide-treated callus, Anaerobic treated  
 seedlings, NAA (a-Naphthalene acetic acid)-treated  
 seedlings, Kinetin-treated seedlings, ACPC  
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
 Brassinolide-treated seedlings, ABA (Abscisisic  
 acid)-treated seedlings, GA (Gibberellic acid)-treated  
 seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
 molecules were generated as follows. First-strand cDNA was  
 prepared from oligo-dT selected mRNA by priming with a  
 NotI oligo-dT primer (5'  
 AACTGGAAGATTCGCGCCGCGAGGAATTTTTTTTTTTTTTTT). The  
 resulting DNA:RNA hybrid was treated with RNase H and used  
 as a template for DNA PolI-catalyzed second strand  
 synthesis. After the addition of EcoRI adaptors, the  
 ds-cDNAs were digested with NotI and size-selected. The  
 resulting molecules were directionally cloned into the  
 EcoRI and NotI sites of the pT7T3PAC vector. The library  
 then went through one round of normalization to Cot value  
 of 5 based on the methods of Marcelo Bento Soares (Genome  
 Research 6: 791-806, 1996)."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 618;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 124 CAAATTAAATGGACTATGCGTGCA 147

## RESULT 11

DE026109  
 LOCUS  
 DEFINITION DE026109 642 bp DNA linear GSS 06-APR-2005  
 Branchiostoma floridae DNA, clone: CH302-041F16.F, genomic survey  
 sequence.  
 DE026109  
 VERSION DE026109.1 GI:62269579  
 KEYWORDS GSS.  
 SOURCE Branchiostoma floridae (Florida lancelet)

```

ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
1
REFERENCE
AUTHORS Fujiyama,A., Toyoda,A., Hattori,M. and Sakaki,Y.
TITLE BAC end sequences of CHORI-302 Amphioxus Library
JOURNAL Published Only in Database (2005)
REFERENCE
AUTHORS Fujiyama,A.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2005) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:afujiyamegsc.riken.jp URL:http://hgp.gsc.riken.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
PRIMERS
Sequencing : T7
LIBRARY
Vector : pTARBAC2.1
R.Site 1 : ECoRI
R.Site 2 : ECoRI.
Location/Qualifiers
1. .642
/organism="Branchiostoma floridae"
/mol_type="genomic DNA"
/db_xref="taxon:7739"
/clone="CH302-041F16.F"
/sex="male"
/tissue type="epERM"
/clone lib="CHORI0302 Amphioxus genomic BAC library"
/note="common name:amphioxus"
ORIGIN
Query Match 76.8%; Score 19.2; DB 11; Length 642;
Best Local Similarity 87.5%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCGTGCAA 25
|| ||||| ||||| |||||
Db 211 AAGTTAAAGAAACTATTCGTGCAA 234

RESULT 12
BM348442 673 bp mRNA linear EST 16-JAN-2002
LOCUS MEST290-H12.T3 ISUM5-RN Zea mays cDNA clone MEST290-H12 3', mRNA
DEFINITION
sequence.
ACCESSION BM348442
VERSION BM348442.1 GI:18173054
KEYWORDS EST.
SOURCE
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 673)
Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA
Tel: 515-294-0975
Fax: 515-294-5256
Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
Phred software,
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/>).
Lucy parameters were set to ensure an overall trimmed quality of
97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve
as spacers.
PCR Primers
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES
Location/Qualifiers
1. .673
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST290-H12"
/tissue_type="mixed"
/lab host="DH10B"
/clone lib="ISUM5-RN"
/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65
DAG), Tassel (3-39 cm, 53 and 56 DAG), Husk (73 DAG), Silk,
(0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Immature ear
unpollinated first ear, ear shank, etiolated seedlings,
callus, Cycloheximide-treated callus, Anaerobic treated
seedlings, NAA (a-Naphthalene acetic acid)-treated
seedlings, Kinetin-treated seedlings, ACPG
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscissic
acid)-treated seedlings, GA (Gibberellic acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dt selected mRNA by priming with a
NotI oligo-dT primer (5',
AATCGAAGATTCGCGCGCAGCAATTTTTTTTTTTTTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-I-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."
ORIGIN
Query Match 76.8%; Score 19.2; DB 3; Length 673;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CAATTAAAGAGACTATTCGTGCA 24
||||| ||||| ||||| |||||
Db 129 CAATTAAAGACTATTCGTGCA 152

RESULT 13
CL680700 765 bp DNA linear GSS 09-JUL-2004
LOCUS PRI012a.D11.2 - PRI012a.BR (765) Mixed stage fosmid library of P.
DEFINITION pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL680700
VERSION CL680700.1 GI:50187660
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 765)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppaDB: an AcedB database for the nematode satellite organism
Pristionchus pacificus

```



JOURNAL  
PUBMED  
COMMENT

Nucleic Acids Res. 32 (1), D421-D422 (2004)  
14681447  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.

FEATURES  
source

Location/Qualifiers  
1..765  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 765;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATTAAAGAGACTATTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 698 AATTATTGAGATATTCGTGCAA 721

RESULT 14

CNS02FKY 785 bp DNA linear GSS 01-SEP-2000  
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
DEFINITION 132H08 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL195163  
VERSION AL195163.1 GI:7833269  
KEYWORDS GSS; genome survey sequence.

SOURCE  
ORGANISM

Tetraodon nigroviridis  
Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS

1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence

JOURNAL  
PUBMED  
REFERENCE

Nat. Genet. 25 (2), 235-238 (2000)  
10835645

REFERENCE  
AUTHORS

2 Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,  
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

JOURNAL  
PUBMED  
REFERENCE

Genome Res. 10 (7), 939-949 (2000)  
10899143

REFERENCE  
AUTHORS

3 (bases 1 to 785)  
Genoscope.

TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- web : www.genoscope.cns.fr)

COMMENT

This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at

FEATURES  
source

Location/Qualifiers  
1..785  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone\_lib="132H08"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG132DD04SP1  
end : PUC-Ori"

ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 785;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATTAAAGAGACTATTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 111 AATTAAAGAGACCAATTGTGCAA 134

RESULT 15  
CRI46965/c

LOCUS CRI46965 804 bp DNA linear GSS 06-JUL-2004  
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and  
chromosome engineering clone MHP345c23, genomic survey sequence.

ACCESSION CRI46965  
VERSION CRI46965.1 GI:49901680  
KEYWORDS GSS; genome survey sequence; MICER.

SOURCE  
ORGANISM

Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 804)  
Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,  
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,  
Rogers,J. and Bradley,A.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. http://www.sanger.ac.uk/MICER

FEATURES  
source

1..804  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="MHP345c23"  
/clone\_lib="MHP3"

ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 804;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTGCA 24  
||||| ||||| ||||| ||||| |||||  
Db 188 CAATCAAGAGATTATTCGACGA 165

RESULT 16  
BZ784243/c

LOCUS BZ784243 810 bp DNA linear GSS 17-MAR-2003  
DEFINITION PUGAT94TB\_ZM\_0\_6\_1.0 KB Zea mays genomic clone ZM5BTa325P20,  
genomic survey sequence.

ACCESSION BZ784243  
VERSION BZ784243.1 GI:28977840  
KEYWORDS GSS.

SOURCE  
ORGANISM

Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 810)

```

AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
              Bennett,J.
TITLE        Maize Genomics Consortium
JOURNAL      Unpublished (2003)
COMMENT      Other GSSs: PUGAT94TD
              Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
  source
    1..810
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBra325P20"
      /clone_lib="ZM_0.6_1.0_KB"
      /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 810;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24
|||||
DB 471 CAAATTAAGAGACTATTCGTGCA 448

RESULT 17
LOCUS      CR173895/c
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
              chromosome engineering clone MHPPI83a09, genomic survey sequence.
ACCESSION  CR173895
VERSION    CR173895.1 GI:49952744
KEYWORDS   GSS; genome survey sequence; MICER.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE   1 (bases 1 to 812)
AUTHORS    Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
              Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
              Rogers,J. and Bradley,A.
TITLE      Direct Submission
JOURNAL    Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES   Location/Qualifiers
            source
              1..812
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /clone="MHPPI83a09"
                /clone_lib="MHPPI"

ORIGIN
Query Match      76.8%; Score 19.2; DB 11; Length 812;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24
|||||
DB 809 CAAATCAAGAGATTATTCGAGCA 786

AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
              Bennett,J.
TITLE        Maize Genomics Consortium
JOURNAL      Unpublished (2003)
COMMENT      Other GSSs: PUGAT94TD
              Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
  source
    1..810
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBra325P20"
      /clone_lib="ZM_0.6_1.0_KB"
      /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 810;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24
|||||
DB 471 CAAATTAAGAGACTATTCGTGCA 448

RESULT 17
LOCUS      CR173895/c
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
              chromosome engineering clone MHPPI83a09, genomic survey sequence.
ACCESSION  CR173895
VERSION    CR173895.1 GI:49952744
KEYWORDS   GSS; genome survey sequence; MICER.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE   1 (bases 1 to 812)
AUTHORS    Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
              Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
              Rogers,J. and Bradley,A.
TITLE      Direct Submission
JOURNAL    Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES   Location/Qualifiers
            source
              1..812
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /clone="MHPPI83a09"
                /clone_lib="MHPPI"

ORIGIN
Query Match      76.8%; Score 19.2; DB 11; Length 812;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24
|||||
DB 809 CAAATCAAGAGATTATTCGAGCA 786

AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
              Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE        Consortium for Maize Genomics
JOURNAL      Unpublished (2002)
COMMENT      Other_GSSs: OG0FUI3TV
              Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

RESULT 18
LOCUS      CG348543/c
DEFINITION OG0FUI3TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0708D01,
              genomic survey sequence.
ACCESSION  CG348543
VERSION    CG348543.1 GI:34265809
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
REFERENCE   1 (bases 1 to 815)
AUTHORS    Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
              Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE      Consortium for Maize Genomics
JOURNAL    Unpublished (2002)
COMMENT      Other_GSSs: OG0FUI3TH
              Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: methylation filtered.
      Location/Qualifiers
        1..815
          /organism="Zea mays"
          /mol_type="genomic DNA"
          /strain="B73"
          /db_xref="taxon:4577"
          /clone="ZMMBMA0708D01"
          /clone_lib="ZM_0.7_1.5_KB"
          /note="Vector: pBC5K; Site 1: HincII; 0.7-1.5 kb
          methylation filtered genomic DNA library"

ORIGIN
Query Match      76.8%; Score 19.2; DB 10; Length 815;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24
|||||
DB 753 CAAATTAAGAGACTATTCGTGCA 730

RESULT 19
LOCUS      CG348533
DEFINITION OG0FUI3TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0708D01,
              genomic survey sequence.
ACCESSION  CG348533
VERSION    CG348533.1 GI:34265799
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
REFERENCE   1 (bases 1 to 816)
AUTHORS    Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
              Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE      Consortium for Maize Genomics
JOURNAL    Unpublished (2002)
COMMENT      Other_GSSs: OG0FUI3TV
              Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

```

Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: methylation filtered.  
Location/Qualifiers

## FEATURES

source

1. .816  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0708D01"  
/clone\_lib="ZM 0.7 1.5 kb"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 816;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24  
|||||  
Db 63 CAAATTAAATGGACTATCGGTGCA 86

## RESULT 20

CC681145 872 bp DNA linear GSS 19-JUN-2003  
LOCUS OGUFK42TV ZM 0.7 1.5 kb Zea mays genomic clone ZMMBMA0428G11,  
DEFINITION genomic survey sequence.

ACCESSION CC681145  
VERSION CC681145.1 GI:32085921  
KEYWORDS GSS.

## SOURCE

Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## TITLE

Consortium for Maize Genomics

## JOURNAL

Unpublished (2002)

## COMMENT

Other GSSs: OGUFK42TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: methylation filtered.

## FEATURES

source

1. .872  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0428G11"  
/clone\_lib="ZM 0.7 1.5 kb"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 872;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24  
|||||  
Db 766 CAAATTAAATGGACTATTCGTGCA 789

## RESULT 21

BG479888

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BG479888 895 bp mRNA linear EST 21-MAR-2001  
602527434Fl NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:4650945 5',  
mRNA sequence.  
BG479888  
BG479888.1 GI:13412167  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 895)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1433 row: C column: 10  
High quality sequence stop: 771.  
Location/Qualifiers

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

1. .895  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4650945"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 21"  
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 895;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24

|||||

Db 843 CAAATCAAGTGCCTATTCGTGCA 866

## RESULT 22

CC681138/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CC681138 925 bp DNA linear GSS 19-JUN-2003  
OGUFK42TH ZM 0.7 1.5 kb Zea mays genomic clone ZMMBMA0428G11,  
genomic survey sequence.  
CC681138  
CC681138.1 GI:32085914  
GSS.

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 925)

## REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,

TITLE Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
JOURNAL Consortium for Maize Genomics  
COMMENT Unpublished (2002)  
Other\_GSSs: OGUFK42TV  
Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: methylation filtered.

Location/Qualifiers

#### FEATURES

1..925  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0428G11"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 925;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24  
|||||  
Db 572 CAAATTAAGACTATCGGTGCA 549

#### RESULT 23

CC686212/c  
LOCUS OGULK68TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0468K15,  
DEFINITION genomic survey sequence.

ACCESSION CC686212.1 GI:32090988  
VERSION  
KEYWORDS GSS.  
SOURCE Zea mays

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 991)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics  
Unpublished (2002)

Other\_GSSs: OGULK68TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: methylation filtered.

Location/Qualifiers

#### FEATURES

1..991  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0468K15"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 991;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24  
|||||  
Db 327 CAAATTAAGACTATCGGTGCA 304

#### RESULT 24

CG033706/c  
LOCUS PUJFW70TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTa0659K19,  
DEFINITION genomic survey sequence.

ACCESSION CG033706  
VERSION CG033706.1 GI:33905862  
KEYWORDS GSS.  
SOURCE Zea mays

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 444)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
Bennerzen,J.

Maize Genomics Consortium  
Unpublished (2003)

Other\_GSSs: PUJFW70TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

#### FEATURES

source  
1..444  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBTa0659K19"  
/clone\_lib="ZM\_0.6\_1.0\_KB"  
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cot selected genomic DNA library"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 444;  
Best Local Similarity 90.9%; Pred. No. 8.9e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTCGTGCA 24  
|||||  
Db 437 AATTAAGAGACTATGTGAA 416

#### RESULT 25

AQ989224/c  
LOCUS Gm\_ISB001\_053\_H20F ISU Soybean BAC Library (pBelBAC11 HindIII)  
DEFINITION Glycine max genomic clone Gm\_ISB001\_053\_H20;; genomic survey  
sequence.

ACCESSION AQ989224  
VERSION AQ989224.1 GI:6937170  
KEYWORDS GSS.  
SOURCE Glycine max (soybean)

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE  
 AUTHORS Marek,L.F., Paz,M., Darnielle,L., Hanson,N. and Shoemaker,R.C.  
 TITLE BAC End sequences from a soybean genomic library (ISU)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Shoemaker Randy C  
 Agronomy Department  
 Iowa State University  
 Ames, IA 50011-1010, USA  
 Tel: 515 294 1205  
 Fax: 515 294 2299  
 Email: rcsshoe@iastate.edu  
 This BAC identified by SSR Satt596. For more information, see  
 Soybase at:  
<http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase>.  
 Please see as an authority for the mapping/naming: Cregan P.B., T.  
 Jarvik, A.L., Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya,  
 T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An  
 integrated genetic linkage map of the soybean genome. Crop Sci.  
 39: (in press)  
 Seq primer: M13F  
 Class: BAC ends

#### FEATURES

source  
 1..460  
 /organism="Glycine max"  
 /mol\_type="genomic DNA"  
 /cultivar="Williams82"  
 /db\_xref="taxon:3847"  
 /clone="Gm15B001\_053\_H20;"  
 /tissue\_type="primary leaves"  
 /dev\_stage="cotyledon stage"  
 /clone\_lib="ISU Soybean BAC Library (pBelobAC11 HindIII)"  
 /notes="Vector: pBelobAC11; The ISU BAC library (Marek and  
 Shoemaker, Genome 40:420, 1997) was constructed using the  
 HindIII site in pBelobAC11. The library consists of  
 approximately 40,000 clones with an average insert size of  
 150 Kb, equal to 4 haploid genome equivalents. Screening  
 of the library is done by PCR amplification of DNA pools."

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 9; Length 460;  
 Best Local Similarity 90.9%; Pred. No. 8.9e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 AATTAAGAGACTATTCGTGC 23  
 Db 297 AATTAAGAGACAATTAGTGC 276

RESULT 26  
 AQ233842  
 LOCUS 612 bp DNA linear GSS 29-SEP-1998  
 DEFINITION HS 2057 A1 B09 T7 CIT Approved Human Genomic Sperm Library D Homo  
 sapiens genomic clone Plate=2057 Col=17 Row=C, genomic survey  
 sequence.  
 ACCESSION AQ233842  
 VERSION AQ233842.1 GI:3662449  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 1 (bases 1 to 612)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 10449764  
 CONTACT: Mahairas GG, Wallace JC, Hood L  
 High throughput Sequencing Center

University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence tagged Connector  
 Plate: 2057 row: C column: 17  
 Class: BAC ends  
 High quality sequence stop: 612.

#### FEATURES

source  
 1..612  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=2057 Col=17 Row=C"  
 /sex="male"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
 E-Coli DH10B"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 9; Length 612;  
 Best Local Similarity 90.9%; Pred. No. 9e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 3 AATTAAGAGACTATTCGTGCA 24  
 Db 424 AATTAAGATACATTCATGCA 445

#### RESULT 27

CV709364  
 LOCUS 148 bp mRNA linear EST 03-NOV-2004  
 DEFINITION UCRPT01\_0012M03 f Poncirus trifoliata CTV-challenged cDNA library -  
 AG12 Poncirus trifoliata cDNA clone PT\_6Ea0012M03, mRNA sequence.  
 ACCESSION CV709364  
 VERSION CV709364.1 GI:55291732  
 KEYWORDS EST.  
 SOURCE Poncirus trifoliata  
 ORGANISM Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Sapindales; Rutaceae; Poncirus.  
 1 (bases 1 to 148)  
 Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,  
 Wanamaker,S., Kim,H.R., Kudrna,D. and Stum,D., Wissotski,M.,  
 Wing,R.  
 Development of EST Resources and New Genetic Markers for California  
 Citrus - Poncirus trifoliata CTV-challenged phloem - AG12

#### REFERENCE

AUTHORS Contact: Mikeal Roose  
 TITLE Department of Botany & Plant Sciences, University of California  
 JOURNAL Department of Botany & Plant Sciences, University of California  
 COMMENT Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124, USA  
 Tel: 9097874137  
 Fax: 9097874437  
 Email: mikeal.roose@ucr.edu  
 Seq primer: T7.

#### FEATURES

source  
 1..148  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"  
 /cultivar="Pomeroy Op"  
 /db\_xref="taxon:37690"  
 /clone="PT\_6Ea0012M03"  
 /tissue\_type="Phloem"  
 /dev\_stage="10 - 30 cm shoots"  
 /lab\_host="E. coli TJCl21"  
 /clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
 library - AG12"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
 greenhouse at University of California, Riverside. The  
 scion was a open-pollinated (very probably selfed)

/strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0047B07"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 [gi|4732114|gb|AF129072.1|, a copy-number inducible derivative of plasmid R1]. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match	74.4%;	Score 18.6;	DB 9;	Length 206;
Best Local Similarity	84.0%;	Pred. No. 1.1e+03;		
Matches 21;	Conservative	0;	Mismatches 4;	Indels

QY I CAAATTAAGAGACIATTCGTGCA 23

132 CAAATTAAAAACACTATTTTGCAA 156

62 JUL 3 1964

DE043516	537 bp	DNA	linear	GSS 25-MAY-2005
Orvzias latipes	DNA	clone: olal-006G19.F.	genomic survey sequence.	

REVISION  
VERSION  
AZ794021.1  
GI:12939565

**KEYWORDS:** GSS.

ORGANISM *Mus musculus*

Eukaryota, Metazoa, Mammalia: Eutheria: Euarchont

REFERENCE 1 (pages 1 to 206)  
SCUDOGNATHI; MARIODEA; MARIODEA

**AUTHORS**  
DUMI, D., ROYAGI, A., BALDE  
TOLAN H. LONGGROVE S. MAB

Islam, H., Longacre, S., Man

Niederhausern, A. and Wright

plasmid inserts

COMMENT Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polyme

**Tel: 801 585 5606**

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Plate: 0047 row: B column: 2

Class: plasmid ends

Class: *Prasmanx class*  
High quality sequence from

1 206  
source  
TELETYPE  
1 206  
TELETYPE/ SOURCE

```
/organelles= Mitochondria
/mag time="genomic"
```

Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: naruse.s.u-tokyo.ac.jp  
 (3) Department of Biological Science,  
 University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: tanarita.s.u-tokyo.ac.jp

## PRIMERS

Sequencing : Forward

## LIBRARY

Vector : pKG145  
 R.Site 1 : SacI  
 L.Site 2 : SacI  
 Location/Qualifiers  
 1. .537  
 /organism="Oryzias latipes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:8090"  
 /clone="Olai-006G19.F"  
 /sex="male"  
 /cell\_type="whole body"  
 /clone\_lib="BAC end sequences of Olai Oryzias latipes library"

FEATURES  
source

## ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 537;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 203 CAAATTACAGAGAAATACGTGCAA 227  
 |||||

## RESULT 30

BZ615728 569 bp DNA linear GSS 16-JAN-2003  
 LOCUS 1954e11.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone  
 DEFINITION 1954e11 5', genomic survey sequence.

ACCESSION BZ615728  
 VERSION BZ615728.1 GI:27766283  
 KEYWORDS GSS.

## SOURCE

ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS Rabinowicz,P.D., O'Shaughnessy,A.L., Ballija,V., Dedhia,N.,  
 Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,  
 Zutavern,T., McCombie,W.R. and Martienssen,R.A.  
 TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)  
 JOURNAL Unpublished (2002)

COMMENT Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874

Email: mcombie@cshl.org  
 Plate: 1954 row: e column: 11  
 Seq primer: -21M13UnivRev  
 Class: shotgun

High quality sequence stop: 569.

Location/Qualifiers

1. .569  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultiivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="1954e11"

FEATURES  
source

/lab\_host="DH5a"  
 /clone\_lib="WGS-ZmaysF (DH5a methyl filtered)"  
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
 The vector was digested with XbaI and one nucleotide was  
 added by fill in the recessive 3' end. The genomic DNA  
 was nebulized, end repaired, adaptor ligated and size  
 fractionated using sephadex. The resulting fragments were  
 between 0.8 and 3 kb and were cloned into the vector (.x/y  
 reads in M13mpl9, .b/g reads in pUC19). The same ligation  
 was transformed into DH5a."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 569;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 100 CAAATGAACAGACTACTCTGTCAA 124  
 |||||

## RESULT 31

CB092313 586 bp mRNA linear EST 27-JAN-2003  
 LOCUS hf03h07.g1 Cycad Leaf Library (NYBG) Cycas rumphii cDNA clone  
 DEFINITION hf03h07, mRNA sequence.

ACCESSION CB092313.1 GI:27916505  
 VERSION CB092313  
 KEYWORDS EST.

## SOURCE

ORGANISM Cycas rumphii  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Cycadophyta; Cycadales; Cycadaceae; Cycas.

## REFERENCE

AUTHORS Brenner,E.D., Katari,M.S., Dedhia,N.N., O'Shaughnessy,A.L.,  
 Ballija,V., Martienssen,R.A., McCombie,W.R., Benfey,P., Coruzzi,G.  
 and Stevenson,D.

Expressed tag sequences from Cycas ovule (NYBG)

## TITLE

JOURNAL Unpublished (2003)

COMMENT Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: hf03 row: h column: 07

Seq primer: -21M13UnivRev

High quality sequence stop: 586.

## FEATURES

Location/Qualifiers

1. .586  
 /organism="Cycas rumphii"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:58031"  
 /clone="hf03h07"  
 /sex="female"

/clone\_lib="Cycad Leaf Library (NYBG)"

/note="Organ: Young leaf; Vector: pBK-CMV; Site 1: Xho I;  
 Site 2: Eco RI; Date: Completed 09/01/2001. Submitted to  
 CSHL 09/05/2001. Sample: Young emergent leaves. From New  
 York Botanical Garden Conservatory accession number 808/59  
 A (collected 03/2001). Library: Made using Strategene's  
 ZAP Express Vector Kit. Library was size fractionated for  
 large inserts."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 6; Length 586;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 485 CAAATGAGAGAGGCTATTCATGCAA 509  
 |||||

## ORIGIN

```

RESULT 32
CX543634
LOCUS
DEFINITION
  UCRPT01_5_008 B05 T7 Poncirus trifoliata CTV-challenged cDNA
  library - UCRPT01-UCR2 Poncirus trifoliata cDNA clone
  UCRPT01_008_T7_B05, mRNA sequence.
ACCESSION
CX543634
VERSION
CX543634.1 GI:57570659
KEYWORDS
EST.
SOURCE
Poncirus trifoliata
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE
1 (bases 1 to 600)
  Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,
  Wanamaker,S., Choi,Y. and Kingan,T.
  Development of EST Resources and New Genetic Markers for California
  Citrus - Poncirus trifoliata CTV-challenged phloem - UCRPT01-UCR2
  Unpublished (2004)
  Contact: Mikeal Roose
  Department of Botany & Plant Sciences, University of California
  Riverside, CA, 92521-0124, USA
  Tel: 9097874137
  Fax: 9097874437
  Email: mikeal.roose@ucr.edu
  Seq primer: T7.
FEATURES
    Location/Qualifiers
      1..600
        /organism="Poncirus trifoliata"
        /mol_type="mRNA"
        /cultivar="Pomeroy Op"
        /db_xref="taxon:37690"
        /clone="UCRPT01_008_T7_B05"
        /tissue_type="bark (with phloem)"
        /dev_stage="10 - 30 cm shoots"
        /lab_host="E. coli TUC121"
        /clone_lib="Poncirus trifoliata CTV-challenged cDNA
        library - UCRPT01-UCR2"
        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
        Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
        greenhouse at University of California, Riverside. The
        scion was an open-pollinated (very probably selfed)
        seedling of Poncirus trifoliata cv Pomeroy that was
        selected as homozygous for the Ctv resistance gene. The
        rootstock was sweet orange infected with citrus tristeza
        virus (CTV) isolate T514 over 1 year before sampling (CTV
        infects sweet orange, but not genotypes carrying the Ctv
        resistance gene. Shoots 10-30 cm long were harvested in
        October 2000, and the green phloem (bark) was removed and
        frozen quickly in dry ice. Total RNA was extracted using
        Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA
        library was made, and 0.5 million primary lambda cDNA
        clones were in vivo excised to give a population of
        Bluescript SK(-) phagemids. All steps to this point were
        performed in the ML Roose lab at the University of
        California, Riverside by X. Ye. Phagemids were plated,
        plaenid DNA purified, cDNA clones archived, and DNA
        sequences determined bi-directionally using an ABI3730 at
        the University of California Riverside Institute of
        Integrative Genome Biology Genomics Core Instrumentation
        Facility, (Choi, Kingan). Chromatogram files were
        downloaded by FTP by Close, then processed by Wanamaker
        (Close lab) using the HarVest pipeline
        (http://harvest.ucr.edu) to remove vector and cloning
        oligo sequences and various contaminants, and to trim to a
        high quality region. Sequences that retained a phred 17
        region of at least 100 bases were assembled, then chimeras
        were removed following manual inspection of assemblies
        (Close, Roose, Wanamaker). Sequences that survived all
        removal steps were submitted to GenBank."

```

```

Query Match      74.4%; Score 18.6; DB 8; Length 600;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 CAAATTAAGAGACTATTTCGTGCA 25
    |||||
Db 114 CAAATTAATACAAATATCCGTCGCA 138
    |||||

RESULT 33
CX644973/C
LOCUS
DEFINITION
  UCRPT02-8F01_g Poncirus trifoliata Roots with Iron Deficiency -
  UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-8F01-L1-1-5-g,
  mRNA sequence.
ACCESSION
CX644973
VERSION
CX644973.1 GI:57879802
KEYWORDS
EST.
SOURCE
Poncirus trifoliata
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE
1 (bases 1 to 604)
  Close,T.J., Roose,M.L., Parker,D.R., Federici,C.F., Mandal,J.,
  Fenton,R.D., Wanamaker,S., Landry,B., Hubert,N., Laforest,M.,
  Landry,J. and Ligonde,A.
  Development of EST Resources and New Genetic Markers for California
  Citrus - Poncirus trifoliata Roots with Iron Deficiency -
  UCRPT02-UCR1
  Unpublished (2005)
  Contact: Timothy J. Close
  Department of Botany & Plant Sciences
  University of California
  Riverside, CA 92521-0124, USA
  Tel: 909-787-3318
  Fax: 909-787-4437
  Email: timothy.close@ucr.edu
  Seq primer: T3.
FEATURES
    Location/Qualifiers
      1..604
        /organism="Poncirus trifoliata"
        /mol_type="mRNA"
        /cultivar="Pomeroy"
        /db_xref="taxon:37690"
        /clone="UCRPT02-8F01-L1-1-5-g"
        /tissue_type="root"
        /dev_stage="seedling"
        /lab_host="E. coli TUC121"
        /clone_lib="Poncirus trifoliata Roots with Iron Deficiency
        - UCRPT02-UCR1"
        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
        Site_1: EcoRI; Site_2: XhoI; Plant materials were prepared
        by Federici (Roose lab) with advice from Parker. Seedlings
        138 days after sowing were about 20-23 cm in height,
        bearing 8-12 leaves. On May 26, 2004, plants were washed
        free of soil with a stream of water, then placed upright
        with the roots submerged in two tanks of nutrient
        solution. The solution was sufficient in all major and
        minor nutrients and buffered with MES at about pH 6.95.
        The plants were maintained in this until June 14, 2004,
        when the solution was changed to one with only 20
        micromolar iron, chelated with EDTA to induce iron
        deficiency. The pH was maintained at 6.99 by sodium
        carbonate/CO2 buffering. This solution was replaced on
        July 6, 2004. Roots from three plants were sampled on
        June 16, June 21, July 1 and July 28, 2004. Roots were
        collected by removing the plant from the nutrient
        solution, blotting off excess moisture with a paper towel,
        then cutting off the top of the plants. Three plants were
        pooled in one aluminum foil packet, and frozen between two

```



sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a trim 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roese, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 604;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25  
|||||  
Db 417 CAAATTAAATCAATATCGTGCAA 393

RESULT 34  
CX644972 608 bp mRNA linear EST 18-JAN-2005  
LOCUS UCRPT02-8F01\_b\_Poncirus trifoliata Roots with Iron Deficiency -  
DEFINITION UCRPT02-UCR1\_Poncirus trifoliata cDNA clone UCRPT02-8F01-L1-1-6.b,  
mRNA sequence.

ACCESSION CX644972.1 GI:57879801  
VERSION CX644972  
KEYWORDS EST.

SOURCE Poncirus trifoliata  
ORGANISM Poncirus trifoliata

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

REFERENCE 1 (bases 1 to 608)  
Close,T.J., Roese,M.L., Parker,D.R., Federici,C.F., Mandal,J.,  
Fenton,R.D., Wanamaker,S., Landry,B., Hubert,N., Laforest,M.,  
Landry,J. and Ligonde,A.

TITLE Development of EST Resources and New Genetic Markers for California  
Citrus - Poncirus trifoliata Roots with Iron Deficiency -

UCRPT02-UCR1

Unpublished (2005)

COMMENT Contact: Timothy J. Close  
Department of Botany & Plant Sciences  
University of California  
Riverside, CA 92521-0124, USA  
Tel: 909-787-3318  
Fax: 909-787-4437

Email: timothy.close@ucr.edu  
Seq primer: T7.

FEATURES  
Source

Location/Qualifiers  
1..608  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy"  
/db\_xref="taxon:37690"  
/clone="UCRPT02-8F01-L1-1-6.b"  
/tissue\_type="root"  
/dev\_stage="seedling"  
/lab\_host="P. coli TUC121"  
/clone\_lib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared by Federici (Roese lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were replaced on July 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a trim 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roese, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 608;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCGCAA 25  
 |||||  
 Db 188 CAAATTAAGAGACTATTCTGTCGCAA 212

RESULT 35  
 CQ340874/c  
 LOCUS CQ340874 666 bp mRNA linear EST 14-JUN-2005  
 DEFINITION tectiformis cDNA clone mtb015n11 3', mRNA sequence.

ACCESSION CQ340874  
 VERSION CQ340874.1 GI:67747923  
 KEYWORDS EST.

SOURCE Molgula tectiformis  
 ORGANISM Molgula tectiformis  
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 Stolidobranchia; Molgulidae; Molgula.

REFERENCE 1 (bases 1 to 666)  
 AUTHORS Gyoja, F., Satou, Y. and Satoh, N.  
 TITLE Expressed genes in Molgula tectiformis  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: Tadaasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6855  
 Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp  
 When you want to obtain this EST clone, please send an e-mail to  
 Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka  
 Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

FEATURES  
 source  
 1..666  
 /organism="Molgula tectiformis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:30286"  
 /clone="mtb015n11"  
 /tissue\_type="whole animal"  
 /dev\_stage="embryo just before hatching"  
 /clone\_lib="Molgula tectiformis unpublished cDNA library"

ORIGIN  
 Query Match 74.4%; Score 18.6; DB 7; Length 666;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCGCAA 25  
 |||||  
 Db 41 CAAATTAAGAGACTATTCTGTCGCAA 17

RESULT 36  
 BI390310/c  
 LOCUS BI390310 670 bp mRNA linear EST 06-AUG-2001  
 DEFINITION pgpic.pk003.o21 Primary Chicken Pituitary/Hypothalamus/Pineal  
 Library Gallus gallus cDNA clone pgpic.pk003.o21 5' similar to no  
 significant hits (pLog(P) > 4), mRNA sequence.

ACCESSION BI390310  
 VERSION BI390310.1 GI:15083592  
 KEYWORDS EST.

SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 670)  
 AUTHORS Porter, T.E. and Cogburn, L.A.  
 TITLE ESTs from Primary Chicken Pituitary/Hypothalamus/Pineal cDNA  
 library USDA/IFAFS Animal Genome Project  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Larry A. Cogburn  
 University of Delaware

Townsend Hall, Newark, DE 19717, USA  
 Tel: 302-831-1335  
 Fax: 302-831-2822  
 Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES  
 source

1..670

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="Commercial broiler chicken"

/db\_xref="taxon:9031"

/clone="pgpic.pk003.o21"

/sex="Male and Female"

/tissue\_type="Pituitary Gland/Hypothalamus/Pineal Gland"

/dev\_stage="Embryonic (dl2,dl4,dl9); post-hatch

(wl,w3,w5,w7,w9)"

/lab\_host="E. Coli EMDH10B"

/clone\_lib="Primary Chicken Pituitary/Hypothalamus/Pineal

Library"

/note="Vector: pCMVSPORT6; Library made from equivalent

pools of total RNA isolated from each tissue at different

ages. Single pass sequencing from 5'-end"

ORIGIN

Query Match 74.4%; Score 18.6; DB 3; Length 670;

Best Local Similarity 84.0%; Pred. No. 1.1e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCGCAA 25

|||||

Db 34 CCAATTAAGAGACTATTCTGTCGCAA 10

|||||

RESULT 37

CC373687/c

LOCUS CC373687 670 bp DNA linear GSS 19-MAY-2003

DEFINITION PUGJN95TB ZM.0.6.1.0 KB Zea mays genomic clone ZMMBTa398P22,

genomic survey sequence.

ACCESSION CC373687

VERSION CC373687.1 GI:30847304

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 670)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Reenick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and

Bennetzen, J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..670

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTa398P22"

/clone\_lib="ZM.0.6.1.0 KB"

/note="Vector: pCE4-TOPO, Site\_1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 670;

Best Local Similarity 84.0%; Pred. No. 1.1e+03;



```

ORGANISM      Molgula tectiformis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
               Stolidobranchia; Molgulidae; Molgula.
REFERENCE     1 (bases 1 to 708)
AUTHORS       Gyoja, F., Satou, Y. and Satoh, N.
TITLE         Expressed genes in Molgula tectiformis
JOURNAL       Unpublished (2005)
COMMENT       Contact: Tadao Shin-i
               Center For Genetic Resource Information
               National Institute of Genetics
               1111 Yata, Mishima, Shizuoka 411-8540, Japan
               Tel: 81-559-81-6856
               Fax: 81-559-81-6855
               Email: tshin@genes.nig.ac.jp
               When you want to obtain this EST clone, please send an e-mail to
               Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
               Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES     Location/Qualifiers
               source
               1..708
               /organism="Molgula tectiformis"
               /mol_type="mRNA"
               /db_xref="taxon:30286"
               /clone="mtgd001c07"
               /tissue_type="gonad"
               /clone_lib="Molgula tectiformis unpublished cDNA library,
               gonad"
ORIGIN
Query Match      74.4%; Score 18.6; DB 7; Length 708;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 25
    ||||| ||||| ||||| |||||
DB 57 CAAATTATAGAGATTATTCTGCA 33
    ||||| ||||| ||||| |||||

RESULT 41
CV509866/c      712 bp mRNA linear EST 05-OCT-2004
LOCUS           kc52a06.y1 Xiphinema index CSEQDA01 Xiphinema index cDNA 5' similar
DEFINITION      to TR-0622295 O62295 Jc8.6B PROTEIN. [2] WP:CE17989 ;, mRNA
sequence.
ACCESSION       CV509866
VERSION         CV509866.1 GI:53807979
KEYWORDS        EST.
SOURCE          Xiphinema index
ORGANISM        Eukaryota; Metazoa; Nematoda; Enoplea; Dorylaimida; Dorylaimina;
               Longidoroidea; Longidoridae; Xiphinema.
REFERENCE       1 (bases 1 to 712)
AUTHORS         McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,
               Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,
               Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,
               Tbagareishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
               Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,
               Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
               McCann, R., Waterston, R. and Wilson, R.
               The Washington Univ. Nematode EST Project, 1999
               Unpublished (1999)
JOURNAL         Contact: Mitreva M
COMMENT         The Washington Univ. Nematode EST Project, 2004
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@wustl.wustl.edu
               Library materials provided by: J.T. Jones Library constructed by:
               Incyte Genomics (for J.T. Jones)
               Seq primer: T7 primer
               High quality sequence stop: 691.
FEATURES     Location/Qualifiers
               source
               1..712

```

```

/organism="Xiphinema index"
/mol_type="mRNA"
/db_xref="taxon:46003"
/sex="mixed"
/dev_stage="mixed"
/lab_host="GC10 (only ligation mix supplied)"
/clone_lib="Xiphinema index CSEQDA01"
/note="Vector: pSport1; Site 1: 5'; Sali; Site 2: 3';
NotI; The library was made from mRNA isolated from total
RNA using oligo dt cellulose. Total RNA was generated from
mixed stage X. index extracted from soil around fig (Ficus
serica) roots. Nematodes were cleaned by passing through a
series of sieves. Pure X. index were obtained by hand
sorting of nematodes. The library was constructed by
Incyte Genomics by cloning cDNA directionally into the
pSport1 vector using Sali and NotI sites. The library was
provided by John T. Jones (jones@scri.sari.ac.uk) at
SCRI, Nematology Department. Sequencing by: Washington
University Genome Sequencing Center, St. Louis, MO."
ORIGIN
Query Match      74.4%; Score 18.6; DB 7; Length 712;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 25
    ||||| ||||| ||||| |||||
DB 505 CAAATTATTGTGACAAATTCGTGCA 481
    ||||| ||||| ||||| |||||

RESULT 42
CV2287609      712 bp DNA linear GSS 01-JUL-2005
LOCUS           cp53e11.r Candida parapsilosis Random Genomic Library Candida
DEFINITION      parapsilosis genomic clone cp53e11, genomic survey sequence.
ACCESSION       CV2287609
VERSION         CV2287609.1 GI:68458919
KEYWORDS        GSS.
SOURCE          Candida parapsilosis
ORGANISM        Candida parapsilosis
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE       1 (bases 1 to 712)
AUTHORS         Logue, M.E., Wong, S., Wolfe, K.H. and Butler, G.
TITLE           A genome sequence survey shows that the pathogenic yeast Candida
               parapsilosis has a defective MTL1 allele at its mating type locus
JOURNAL         Eukaryot. Cell 4 (6), 1009-1017 (2005)
PUBMED          15947193
COMMENT         Contact: Logue M
               Department of Biochemistry, Conway Institute of Biomolecular and
               Biomedical Research
               University College Dublin
               Dublin 4, Ireland
               Tel: +353 1 7166885
               Fax: +353 1 2837211
               Email: mary.e.logue@ucd.ie
               Class: mapped ends.
               Location/Qualifiers
               source
               1..712
               /organism="Candida parapsilosis"
               /mol_type="genomic DNA"
               /strain="CU1B214"
               /db_xref="taxon:5480"
               /clone="cp53e11"
               /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN
Query Match      74.4%; Score 18.6; DB 10; Length 712;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 25
    ||||| ||||| ||||| |||||

```

### Query Match

high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

```

ORIGIN
Query Match          74.4%; Score 18.6; DB 7; Length 727;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCAA 25
    |||||
Db 156 CAAATTAACATATATCGTGCAA 180
    |||||

RESULT 45
CJ344248/c 731 bp mRNA linear EST 14-JUN-2005
LOCUS CJ344248 Molgula tectiformis unpublished cDNA library Molgula
DEFINITION tectiformis cDNA clone mtbh025ml2 3', mRNA sequence.
ACCESSION CJ344248
VERSION CJ344248.1 GI:67751297
KEYWORDS EST.
SOURCE Molgula tectiformis
ORGANISM Molgula tectiformis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Molgulidae; Molgula.
REFERENCE 1 (bases 1 to 731)
AUTHORS Gyoja,F., Satou,Y. and Satoh,N.
TITLE Expressed genes in Molgula tectiformis
JOURNAL Unpublished (2005)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
When you want to obtain this EST clone, please send an e-mail to
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES
source
1..731
/organism="Molgula tectiformis"
/mol_type="mRNA"
/db_xref="taxon:30286"
/clone="mtbh025ml2"
/tissue_type="whole animal"
/dev_stage="embryo just before hatching"
/clone_lib="Molgula tectiformis unpublished cDNA library"

ORIGIN
Query Match          74.4%; Score 18.6; DB 7; Length 731;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCAA 25
    |||||
Db 44 CAAATTATAGAGATTATTCTGCAA 20
    |||||

RESULT 46
CJ357933/c 732 bp mRNA linear EST 15-JUN-2005
LOCUS CJ357933 Molgula tectiformis unpublished cDNA library, cleaving
DEFINITION embryo Molgula tectiformis cDNA clone mtcl006f20 3', mRNA sequence.
ACCESSION CJ357933
VERSION CJ357933.1 GI:67798081
KEYWORDS EST.
SOURCE Molgula tectiformis
ORGANISM Molgula tectiformis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Molgulidae; Molgula.
REFERENCE 1 (bases 1 to 732)
AUTHORS Gyoja,F., Satou,Y. and Satoh,N.

```

```

TITLE Expressed genes in Molgula tectiformis
JOURNAL Unpublished (2005)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
When you want to obtain this EST clone, please send an e-mail to
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES
source
1..732
/organism="Molgula tectiformis"
/mol_type="mRNA"
/db_xref="taxon:30286"
/clone="mtcl006f20"
/tissue_type="whole animal"
/dev_stage="cleaving embryo"
/clone_lib="Molgula tectiformis unpublished cDNA library,
cleaving embryo"

ORIGIN
Query Match          74.4%; Score 18.6; DB 7; Length 732;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCAA 25
    |||||
Db 59 CAAATTATAGAGATTATTCTGCAA 35
    |||||

RESULT 47
CJ641978 733 bp mRNA linear EST 18-JAN-2005
LOCUS CJ641978 UCRPT02-57B01-b Poncirus trifoliata Roots with Iron Deficiency -
DEFINITION UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-57B01-D2-1-6.b,
mRNA sequence.
ACCESSION CJ641978
VERSION CJ641978.1 GI:57876807
KEYWORDS EST.
SOURCE Poncirus trifoliata
ORGANISM Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE 1 (bases 1 to 733)
AUTHORS Close,T.J., Roose,M.L., Parker,D.R., Federici,C.F., Mandal,J.,
Penton,R.D., Wanamaker,S., Landry,B., Hubert,N., Laforest,M.,
Landry,J. and Ligonde,A.
Development of EST Resources and New Genetic Markers for California
Citrus - Poncirus trifoliata Roots with Iron Deficiency -
UCRPT02-UCR1
JOURNAL Unpublished (2005)
COMMENT Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T7.
location/Qualifiers
1..733
/organism="Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy"
/db_xref="taxon:37690"
/clone="UCRPT02-57B01-D2-1-6.b"
/tissue_type="root"
/dev_stage="seedling"
/lab_host="E. coli TJC121"

```

/clone\_lib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared by Federici (Roose Lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on June 16, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using Trizol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 733;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGACACTATTCGTGCAA 25  
 |||||  
 Db 153 CAATTAATAACAATATTCGTGCAA 177  
 |||||

## RESULT 48

CZ279576  
 LOCUS 740 bp DNA linear GSS 01-JUL-2005  
 DEFINITION cp06e05.f Candida parapsilosis Random Genomic Library Candida parapsilosis genomic clone cp06e05, genomic survey sequence.  
 ACCESSION CZ279576

VERSION CZ279576.1 GI:68450778  
 KEYWORDS GSS.

## SOURCE

ORGANISM Candida parapsilosis  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.

## REFERENCE

1 (bases 1 to 740)  
 Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.

## AUTHORS

A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective MTL1 allele at its mating type locus

## TITLE

Eukaryot. Cell 4 (6), 1009-1017 (2005)

## JOURNAL

PUBMED 15947133

## COMMENT

Contact: Logue M  
 Department of Biochemistry, Conway Institute of Biomolecular and Biomedical Research  
 University College Dublin  
 Dublin 4, Ireland  
 Tel: +353 1 7166885  
 Fax: +353 1 2837211  
 Email: mary.e.logue@ucd.ie  
 Class: plasmid ends.

## FEATURES

source

Location/Qualifiers  
 1..740  
 /organism="Candida parapsilosis"  
 /mol\_type="genomic DNA"  
 /strain="CU1B214"  
 /db\_xref="taxon:5480"  
 /clone="cp06e05"  
 /clone\_lib="Candida parapsilosis Random Genomic Library"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 740;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGACACTATTCGTGCAA 25  
 |||||  
 Db 582 CCAATTGAAAGACTATTCGAGCAA 606  
 |||||

## RESULT 49

CJ361428/c  
 LOCUS

## DEFINITION

CJ361428 Molgula tectiformis unpublished cDNA library, cleaving embryo Molgula tectiformis cDNA clone mtc1018f13 3', mRNA sequence.

## ACCESSION

CJ361428 GI:67801576

## VERSION

EST.

## KEYWORDS

ORGANISM Molgula tectiformis  
 Molgula tectiformis  
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Stolidobranchia; Molgulidae; Molgula.

## REFERENCE

1 (bases 1 to 742)  
 Gyoja,F., Satou,Y. and Satoh,N.

## AUTHORS

Expressed genes in Molgula tectiformis

## TITLE

Unpublished (2005)

## JOURNAL

COMMENT Contact: Tadau Shin-i

## COMMENT

Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp

When you want to obtain this EST clone, please send an e-mail to Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

## FEATURES

source

Location/Qualifiers  
 1..742  
 /organism="Molgula tectiformis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:30286"  
 /clone="mtc1018f13"  
 /tissue\_type="whole animal"





Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.choi.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering information.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 457 row: M column: 24  
 Seq primer: SP6  
 Class: BAC ends.

#### FEATURES

source  
 Location/Qualifiers  
 1..758  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SENHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-457M24"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
 Pieter de Jong"

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 758;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCAA 25  
 |||||  
 Db 8 CAAATTAAAGAGACTATTCTGTGAAA 32

#### RESULT 53

CV708436 764 bp mRNA linear EST 03-NOV-2004  
 LOCUS UCRPT01\_0011F05 f Poncirus trifoliata CTV-challenged cDNA library -  
 DEFINITION AG12 Poncirus trifoliata cDNA clone PT\_6Ea0011F05, mRNA sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

Poncirus trifoliata  
 Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

#### REFERENCE

#### AUTHORS

1 (bases 1 to 764)  
 Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,  
 Wanmaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wissotski, M.,  
 Wing, R.

#### TITLE

Development of EST Resources and New Genetic Markers for California  
 Citrus - Poncirus trifoliata CTV-challenged phloem - AG12  
 Unpublished (2004)

#### JOURNAL

#### COMMENT

Contact: Mikeal Roose  
 Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124, USA  
 Tel: 9097874137  
 Fax: 9097874437  
 Email: mikeal.roose@ucr.edu

#### Seq primer: T7

#### Location/Qualifiers

1..764  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"  
 /cultivar="Pomeroy Op"  
 /db\_xref="taxon:37690"  
 /clone="PT\_6Ea0011F05"  
 /tissue\_type="Phloem"  
 /dev\_stage="10 - 30 cm shoots"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
 library - AG12"

#### FEATURES

#### source

Location/Qualifiers  
 1..788  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"

#### FEATURES

#### source

#### Location/Qualifiers

#### seq primer: T7

#### Location/Qualifiers

#### seq primer: T7

#### Location/Qualifiers

#### seq primer: T7

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
 greenhouse at University of California, Riverside. The  
 action was an open-pollinated (very probably selfed)  
 seedling of Poncirus trifoliata cv Pomeroy that was  
 selected as homozygous for the Ctv resistance gene. The  
 rootstock was sweet orange infected with citrus tristeza  
 virus (CTV) isolate TS14 over 1 year before sampling the CTV  
 infects sweet orange, but not genotypes carrying the Ctv  
 resistance gene). Shoots 10-30 cm long were harvested in  
 October 2000, and the green phloem (bark) was removed and  
 frozen quickly in dry ice. Total RNA was extracted using  
 TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
 library was made, and 0.5 million primary lambda cDNA  
 clones were in vivo excised to give a population of  
 pBluescript SK(-) phagemids. All steps to this point were  
 performed in the M. Roose lab at the University of  
 California, Riverside by X. Ye. Phagemids were plated,  
 plasmid DNA purified, cDNA clones archived, and DNA  
 sequences determined bi-directionally using an ABI3730 at  
 the Arizona Genomics Institute, University of Arizona  
 (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
 were downloaded to UC Riverside (Close), then processed at  
 UC Riverside (Wanmaker) using the HarVest pipeline  
 (http://harvest.ucr.edu) to remove vector and cloning  
 oligo sequences and various contaminants, and to trim to a  
 high quality region. Sequences that retained a phred 17  
 region of at least 100 bases were deposited to GenBank."

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 764;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCAA 25  
 |||||  
 Db 115 CAAATTAAATACATATCTGTGCAA 139

#### RESULT 54

#### LOCUS

#### DEFINITION

CX644101 788 bp mRNA linear EST 18-JAN-2005  
 UCRPT02\_71G08 b Poncirus trifoliata Roots with Iron Deficiency -  
 UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-71G08-N16-6.b,  
 mRNA sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### Poncirus trifoliata

#### Poncirus trifoliata

#### Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

#### Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

#### rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

#### REFERENCE

#### AUTHORS

#### 1 (bases 1 to 788)

#### Fenton, R.D., Wanmaker, S., Landry, B., Hubert, N., Laforest, M.,

#### Landry, J. and Ligonde, A.

#### Development of EST Resources and New Genetic Markers for California

#### Citrus - Poncirus trifoliata Roots with Iron Deficiency -

#### UCRPT02-UCR1

#### Unpublished (2005)

#### Contact: Timothy J. Close

#### Department of Botany & Plant Sciences

#### University of California

#### Riverside, CA 92521-0124, USA

#### Tel: 909-787-3318

#### Fax: 909-787-4437

#### Email: timothy.close@ucr.edu

#### Seq primer: T7

#### Location/Qualifiers

#### 1..788

#### /organism="Poncirus trifoliata"

#### /mol\_type="mRNA"

/cultivar="Pomeroy"  
 /db\_xref="taxon:37690"  
 /clone="UCRPT02-71G08-N16-6.b"  
 /tissue\_type="root"  
 /dev\_stage="seedling"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared by Federici (Roose lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanmaker (Close lab) using the HarVest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wanmaker). Sequences that survived all removal steps were submitted to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 788;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTCAA 25  
 |||||  
 Db 137 CAAATTAATACAAATTCGTCAA 161  
 |||||

## RESULT 55

CV706539

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

Location/Qualifiers

1..793

/organism="Poncirus trifoliata"

/mol\_type="mRNA"

/cultivar="Pomeroy OP"

/db\_xref="taxon:37690"

/clone="PT 6EA0008H19"

/tissue\_type="Phloem"

/dev\_stage="10 - 30 cm shoots"

/lab\_host="E. coli TJC121"

/clone\_lib="Poncirus trifoliata CTV-challenged cDNA library - AG12"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The scion was an open-pollinated (very probably selfed) seedling of *Poncirus trifoliata* cv Pomeroy that was selected as homozygous for the CTV resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate TS14 over 1 year before sampling (CTV infects sweet orange, but not genotypes carrying the CTV resistance gene). Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files were downloaded to UC Riverside (Close), then processed at UC Riverside (Wanmaker) using the HarVest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 793;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;



**AUTHORS** Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,  
Wanamaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wissotski, M.,  
Wing, R.  
**TITLE** Development of EST Resources and New Genetic Markers for California  
Citrus - Poncirus trifoliata CTV-challenged phloem - AG12  
**JOURNAL** Unpublished (2004)  
**COMMENT** Contact: Mikeal Roose  
Department of Botany & Plant Sciences, University of California  
Riverside, CA, 92521-0124, USA  
Tel: 9097874137  
Fax: 9097874437  
Email: mikeal.roose@ucr.edu  
Seq primer: T7

**FEATURES** Location/Qualifiers  
source  
1..810  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy OP"  
/db\_xref="taxon:37690"  
/clone="PT\_6Ea0006007"  
/tissue\_type="Phloem"  
/dev\_stage="10 - 30 cm shoots"  
/lab\_host="E. coli TJC121"  
/clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
library - AG12"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
greenhouse at University of California, Riverside. The  
scion was an open-pollinated (very probably selfed)  
seedling of Poncirus trifoliata cv Pomeroy that was  
selected as homozygous for the Ctv resistance gene. The  
rootstock was sweet orange infected with citrus tristeza  
virus (CTV) isolate T514 over 1 year before sampling (CTV  
infests sweet orange, but not genotypes carrying the Ctv  
resistance gene). Shoots 10-30 cm long were harvested in  
October 2000, and the green phloem (bark) was removed and  
frozen quickly in dry ice. Total RNA was extracted using  
TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
library was made, and 0.5 million primary lambda cDNA  
clones were in vivo excised to give a population of  
pBluescript SK(-) phagemids. All steps to this point were  
performed in the ML Roose lab at the University of  
California, Riverside by X. Ye. Phagemids were plated,  
plasmid DNA purified, cDNA clones archived, and DNA  
sequences determined bi-directionally using an ABI3730 at  
the Arizona Genomics Institute, University of Arizona  
(Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
were downloaded to UC Riverside (Close), then processed at  
UC Riverside (Wanamaker) using the HarVest pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that retained a phred 17  
region of at least 100 bases were deposited to GenBank."

**ORIGIN**  
Query Match 74.4%; Score 18.6; DB 7; Length 810;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGTCGA 25  
|||||  
DB 190 CAATTAAGAGACTATTCTGTCGA 214  
|||||  
RESULT 59  
CV708661 816 bp mRNA linear EST 03-NOV-2004  
LOCUS UCRPT01\_0011K18 f Poncirus trifoliata CTV-challenged cDNA library -  
DEFINITION AG12 Poncirus trifoliata cDNA clone PT\_6Ea0011K18, mRNA sequence.  
ACCESSION CV708661  
VERSION CV708661.1 GI:55291029  
KEYWORDS EST.  
SOURCE Poncirus trifoliata

**ORGANISM**  
**REFERENCE**  
**AUTHORS**

**TITLE**  
**JOURNAL**  
**COMMENT**

**FEATURES**  
source

1..816  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy OP"  
/db\_xref="taxon:37690"  
/clone="PT\_6Ea0011K18"  
/tissue\_type="Phloem"  
/dev\_stage="10 - 30 cm shoots"  
/lab\_host="E. coli TJC121"  
/clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
library - AG12"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
greenhouse at University of California, Riverside. The  
scion was an open-pollinated (very probably selfed)  
seedling of Poncirus trifoliata cv Pomeroy that was  
selected as homozygous for the Ctv resistance gene. The  
rootstock was sweet orange infected with citrus tristeza  
virus (CTV) isolate T514 over 1 year before sampling (CTV  
infests sweet orange, but not genotypes carrying the Ctv  
resistance gene). Shoots 10-30 cm long were harvested in  
October 2000, and the green phloem (bark) was removed and  
frozen quickly in dry ice. Total RNA was extracted using  
TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
library was made, and 0.5 million primary lambda cDNA  
clones were in vivo excised to give a population of  
pBluescript SK(-) phagemids. All steps to this point were  
performed in the ML Roose lab at the University of  
California, Riverside by X. Ye. Phagemids were plated,  
plasmid DNA purified, cDNA clones archived, and DNA  
sequences determined bi-directionally using an ABI3730 at  
the Arizona Genomics Institute, University of Arizona  
(Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
were downloaded to UC Riverside (Close), then processed at  
UC Riverside (Wanamaker) using the HarVest pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that retained a phred 17  
region of at least 100 bases were deposited to GenBank."

**ORIGIN**  
Query Match 74.4%; Score 18.6; DB 7; Length 816;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGTCGA 25  
|||||  
DB 175 CAATTAAGAGACTATTCTGTCGA 199  
|||||

RESULT 60  
CV401436/c  
LOCUS CJ401436  
DEFINITION CJ401436 Molgula tectiformis unpublished cDNA library, gonad

823 bp mRNA linear EST 21-JUN-2005

Molgula tectiformis cDNA clone mtgd015b02 3', mRNA sequence.

ACCESSION CJ401436  
VERSION CJ401436.1 GI:68093391  
KEYWORDS EST.  
SOURCE Molgula tectiformis  
ORGANISM Molgula tectiformis  
REFERENCE Molgula tectiformis  
AUTHORS Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
TITLE Stolidobranchia; Molgulidae; Molgula.  
JOURNAL 1 (bases 1 to 823)  
COMMENT Gyoja, F., Satou, Y. and Satoh, N.  
Expressed genes in Molgula tectiformis  
Unpublished (2005)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp  
When you want to obtain this EST clone, please send an e-mail to  
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka  
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).  
FEATURES  
source  
1..823  
Location/Qualifiers  
/organism="Molgula tectiformis"  
/mol\_type="mRNA"  
/db\_xref="taxon:30286"  
/clone="mtgd015b02"  
/tissue\_type="gonad"  
/clone\_lib="Molgula tectiformis unpublished cDNA library,  
gonad"  
ORIGIN  
Query Match 74.4%; Score 18.6; DB 7; Length 823;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 CAAATTAAAGAGACTATTCTGCGAA 25  
Db 62 CAAATTATAGAGATTATTCTGCAA 38

Search completed: January 28, 2006, 01:14:26  
Job time : 2035.12 secs

***This Page Blank (uspto)***

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 70.2247 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaattaaagagactattctgca 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	77.6	162841	3	US-09-949-016-13733
2	18.2	72.8	600	3	Sequence 3094, Ap
3	17.6	70.4	512	2	US-08-460-739-2
4	17.6	70.4	577	3	US-09-270-767-7869
5	17.6	70.4	577	3	US-09-270-767-23151
6	17.6	70.4	2947	3	US-09-710-279-3805
7	17.6	70.4	3285	3	US-09-710-279-3153
8	17.6	70.4	3730	3	US-09-710-279-3499
9	17.6	70.4	5418	3	US-09-949-016-14328
10	17.6	70.4	30549	3	US-09-134-001C-322
11	17.6	70.4	236474	3	US-09-949-016-13418
12	17.6	70.4	260247	3	US-09-949-016-13358
13	17.6	70.4	1664976	3	US-08-916-421B-1
14	17.6	70.4	1664976	3	US-09-692-570-1
15	17.4	69.6	13201	3	US-09-949-016-14672
16	17.2	68.8	490	3	US-09-513-999C-32487
17	17.6	70.4	601	3	US-09-949-016-140776
18	17.6	70.4	601	3	US-09-949-016-140777
19	17.6	70.4	601	3	US-09-949-016-142111
20	17.6	70.4	747	3	US-08-956-171E-653
21	17.6	70.4	747	3	US-08-781-986A-653
22	17.6	70.4	1140	3	US-09-583-110-1191
23	17.6	70.4	1287	3	US-09-107-433-1469
24	17.6	70.4	1323	3	US-10-081-923-12

25	17	68.0	1378	3	US-10-081-923-14	Sequence 14, Appl
26	17	68.0	1379	3	US-10-081-923-7	Sequence 7, Appl
27	17	68.0	1379	3	US-10-081-923-13	Sequence 13, Appl
28	17	68.0	1379	3	US-10-081-923-16	Sequence 16, Appl
29	17	68.0	1384	3	US-10-081-923-8	Sequence 8, Appl
30	17	68.0	1384	3	US-10-081-923-10	Sequence 10, Appl
31	17	68.0	1385	3	US-10-081-923-9	Sequence 9, Appl
32	17	68.0	1390	3	US-10-081-923-11	Sequence 11, Appl
33	17	68.0	1393	3	US-10-081-923-15	Sequence 15, Appl
34	17	68.0	1426	3	US-09-075-460-1	Sequence 1, Appl
35	17	68.0	1447	3	US-10-081-923-17	Sequence 17, Appl
36	17	68.0	1545	3	US-09-248-796A-46	Sequence 46, Appl
37	17	68.0	1795	2	US-08-579-667-1	Sequence 1, Appl
38	17	68.0	1803	3	US-10-081-923-5	Sequence 5, Appl
39	17	68.0	2114	3	US-10-104-047-1637	Sequence 1637, Ap
40	17	68.0	2811	3	US-09-248-796A-740	Sequence 740, App
41	17	68.0	8195	3	US-08-961-527-94	Sequence 94, Appl
42	17	68.0	44064	3	US-09-949-016-12015	Sequence 12015, A
43	17	68.0	44072	3	US-09-949-016-15757	Sequence 15757, A
44	17	68.0	88758	3	US-09-949-016-13502	Sequence 13502, A
45	17	68.0	96845	3	US-09-949-016-13658	Sequence 13658, A
46	17	68.0	264358	3	US-09-949-016-15725	Sequence 15725, A
47	16.8	67.2	346	3	US-09-325-932A-133	Sequence 133, App
48	16.8	67.2	1653	3	US-09-325-932A-132	Sequence 132, App
49	16.6	66.4	462	3	US-09-798-635A-10	Sequence 10, Appl
50	16.6	66.4	506	3	US-09-621-976-18917	Sequence 18917, A
51	16.6	66.4	526	3	US-09-270-767-4638	Sequence 4638, Ap
52	16.6	66.4	526	3	US-09-270-767-19920	Sequence 19920, A
53	16.6	66.4	601	3	US-09-949-016-34577	Sequence 34577, A
54	16.6	66.4	601	3	US-09-949-016-34578	Sequence 34578, A
55	16.6	66.4	601	3	US-09-949-016-48887	Sequence 48887, A
56	16.6	66.4	601	3	US-09-949-016-48888	Sequence 48888, A
57	16.6	66.4	601	3	US-09-949-016-49059	Sequence 49059, A
58	16.6	66.4	601	3	US-09-949-016-49060	Sequence 49060, A
59	16.6	66.4	601	3	US-09-949-016-49231	Sequence 49231, A
60	16.6	66.4	601	3	US-09-949-016-49232	Sequence 49232, A

ALIGNMENTS

RESULT 1  
US-09-949-016-13733  
; Sequence 13733, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13733  
; LENGTH: 162841  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13733

Query Match 77.6%; Score 19.4; DB 3; Length 162841;  
Best Local Similarity 95.2%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTG 22

Db 12094 AAATTAAGAGACTATTCGTG 12114

```

; ORGANISM: DNA for encoding CFA1 protein
US-08-460-739-2

Query Match 70.4%; Score 17.6; DB 2; Length 512;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
    ||||| ||||| |||||
DB 5 AATTTAAAAAACTATTGGTGCAA 28
    ||||| ||||| |||||

RESULT 4
US-09-270-767-7869
; Sequence 7869, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7869
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-7869

Query Match 70.4%; Score 17.6; DB 3; Length 577;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
    ||||| ||||| |||||
DB 478 AAATTAAGCAACAATTGTGCAA 501
    ||||| ||||| |||||

RESULT 5
US-09-270-767-23151
; Sequence 23151, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23151
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-23151

Query Match 70.4%; Score 17.6; DB 3; Length 577;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
    ||||| ||||| |||||
DB 478 AAATTAAGCAACAATTGTGCAA 501
    ||||| ||||| |||||

RESULT 6
US-09-710-279-3805
; Sequence 3805, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

```



; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3805  
; LENGTH: 2947  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3805

Query Match 70.4%; Score 17.6; DB 3; Length 2947;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1894 CAAATAAAGAGAAATAAGTGCA 1917

RESULT 7  
US-09-710-279-3153  
; Sequence 3153, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3153  
; LENGTH: 3285  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3153

Query Match 70.4%; Score 17.6; DB 3; Length 3285;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1447 CAAATAAAGAGAAATAAGTGCA 1470

RESULT 8  
US-09-710-279-3499  
; Sequence 3499, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3499  
; LENGTH: 3730

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3499

Query Match 70.4%; Score 17.6; DB 3; Length 3730;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1892 CAAATAAAGAGAAATAAGTGCA 1915

RESULT 9  
US-09-949-016-14328  
; Sequence 14328, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14328  
; LENGTH: 5418  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14328

Query Match 70.4%; Score 17.6; DB 3; Length 5418;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCGTGCAA 25  
Db 4733 AAATTAAAGAGATTATTCTAGCAA 4756

RESULT 10  
US-09-134-001C-322  
; Sequence 322, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 322  
; LENGTH: 30549  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-322

Query Match 70.4%; Score 17.6; DB 3; Length 30549;

Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTGCA 24  
||||| ||||| ||||| ||||| |||||  
Db 12031 CAAATTAAGAGAAATAAGTGCA 12054

RESULT 11  
US-09-949-016-13418/c  
; Sequence 13418, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13418  
; LENGTH: 236474  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(236474)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-13418

Query Match 70.4%; Score 17.6; DB 3; Length 236474;

Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 197388 AAATTAAGAGACTATTCGTGTA 197365

RESULT 12  
US-09-949-016-13358  
; Sequence 13358, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13358  
; LENGTH: 260247  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13358

Query Match 70.4%; Score 17.6; DB 3; Length 260247;

Best Local Similarity 83.3%; Pred. No. 2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 152084 AACTTAAGAGATATGTGTCAA 152107

RESULT 13  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.  
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
; Jannaschii  
; Patent No. 6503729  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; CURRENT FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; PRIOR FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84773)..(84773)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84808)..(84808)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84812)..(84812)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98120)..(98120)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98159)..(98159)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98239)..(98239)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98266)..(98266)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98343)..(98343)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (103998)..(103998)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (148948)..(148948)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (163385)..(163385)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191989)..(191989)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191995)..(191995)  
; OTHER INFORMATION: n equals a, t, c, or g

```
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```

```
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

```
Query Match 70.4%; Score 17.6; DB 3; Length 1664976;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 2 AAATTAAAGAGACTATTTCGTGCAA 25
Db 1522836 AAATTAAAGAGAAATTAATGCAA 1522859
```

```
RESULT 14
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
```

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g

Query Match 70.4%; Score 17.6; DB 3; Length 1664976;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25  
Db 1522836 AAATTAAGAGAAATTAATGCAA 1522859

RESULT 15  
US-09-949-016-14672/c  
Sequence 14672, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14672  
LENGTH: 13201  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-14672

Query Match 69.6%; Score 17.4; DB 3; Length 13201;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCG 20  
|||||

Db 5301 AAATTAAGAGACTATTCG 5283

RESULT 16  
US-09-513-999C-32487  
Sequence 32487, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961  
FILE REFERENCE: 59.US2.REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 32487  
LENGTH: 490  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-513-999C-32487

Query Match 68.8%; Score 17.2; DB 3; Length 490;  
Best Local Similarity 86.4%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 377 ATTAAGACTATTAGTGAA 398

RESULT 17  
US-09-949-016-140776  
Sequence 140776, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 140776  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-140776

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 483 CACATTAAGAACTATTGAAGCAA 507

RESULT 18  
US-09-949-016-140777  
Sequence 140777, Application US/09949016  
Patent No. 6812339

;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 140777  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-140777

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAAATTAAGAGACTATTTCGTGCAA 25  
Db 491 CACATTAAAGAACTATTGAAGCAA 515

RESULT 19  
US-09-949-016-142111/c  
; Sequence 142111, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 142111  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-142111

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAAATTAAGAGACTATTTCGTGCAA 25  
Db 125 CAAATTCATGAGCAATTCATGCAA 101

RESULT 20  
US-08-956-171E-653/c  
; Sequence 653, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen

;; Steven C. Barash  
;; Michael R. Fannon  
;; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
;; NUMBER OF SEQUENCES: 5256  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
;; COMPUTER: HP Vectra 486/33  
;; OPERATING SYSTEM: MSDOS version 6.2  
;; SOFTWARE: ASCII Text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/956,171E  
;; FILING DATE: 20-Oct-1997  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/009,861  
;; FILING DATE: January 5, 1996  
;; APPLICATION NUMBER: 08/781,986  
;; FILING DATE: January 3, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mark J. Hyman  
;; REGISTRATION NUMBER: 46,789  
;; REFERENCE/DOCKET NUMBER: PB248P1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (240) 314-1224  
;; TELEFAX: (301) 309-8439  
;; INFORMATION FOR SEQ ID NO: 653:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 747 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 653:  
US-08-956-171E-653

Query Match 68.0%; Score 17; DB 3; Length 747;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTTCGTGCAA 25  
Db 296 CAAATCAATGACAAAATTCGTGCAA 272

RESULT 21  
US-08-781-986A-653/c  
; Sequence 653, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 653:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-653

Query Match 68.0%; Score 17; DB 3; Length 747;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 296 CAAATCAATGACAAATTCGTGCAA 272

RESULT 22
US-09-583-110-1191
; Sequence 1191, Application US/09583110
; Patent No. 6698703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1191
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-583-110-1191

Query Match 68.0%; Score 17; DB 3; Length 1140;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 265 CAAATCAAGGAAACTATCCGTGCGA 289

RESULT 23
US-09-107-433-1469
; Sequence 1469, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
```

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1469:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1287
; SEQUENCE DESCRIPTION: SEQ ID NO: 1469:
; US-09-107-433-1469

Query Match 68.0%; Score 17; DB 3; Length 1287;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 319 CAAATCAAGGAAACTATCCGTGCGA 343
```

```
RESULT 24
US-10-081-923-12
; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pti sequence from isolate no. 2
; US-10-081-923-12

Query Match 68.0%; Score 17; DB 3; Length 1323;
```





## RESULT 32

```
; SEQ ID NO 1
; LENGTH: 1426
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (281)...(1201)
US-09-075-460-1

Query Match      68.0%; Score 17; DB 3; Length 1426;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 455 CAGATTAAAGAGGATATTCCTGAAA 479

RESULT 35
US-10-081-923-17
; Sequence 17, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptA1 sequence from isolate no. 11
US-10-081-923-17

Query Match      68.0%; Score 17; DB 3; Length 1447;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 98 CAGGTTAAAGAAACGATTCGCGCAA 122

RESULT 36
US-09-248-796A-46
; Sequence 46, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 46
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-46

Query Match      68.0%; Score 17; DB 3; Length 1545;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 57 CAAATTAAGTGGCTATATTGAAA 81

RESULT 37
US-08-579-667-1
; Sequence 1, Application US/08579667
; Patent No. 5705624
; GENERAL INFORMATION:
; APPLICANT: Fitzmaurice, Wayne P.
; APPLICANT: Hellmann, Gary M.
; APPLICANT: Grill, Laurence K.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Della-Cioppa, Guy R.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
; TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: 1211 East Morehead Street, PO Drawer 34009
; CITY: Charlotte
; STATE: No. 5705624th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,667
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 627-196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1795 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 362..1591
US-08-579-667-1

Query Match      68.0%; Score 17; DB 2; Length 1795;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 57 CAAATTAAGTGGCTATATTGAAA 81

RESULT 38
US-10-081-923-5
; Sequence 5, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
```

```
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from Oklahoma University M1 strain
; PUBLICATION INFORMATION:
; AUTHORS: Ferretti et al.
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 98
; PAGES: 4658-4663
; DATE: 2001-01-01
; US-10-081-923-5

Query Match      68.0%; Score 17; DB 3; Length 1803;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 265 CAGTTAAAGAAACGATTCGCGCAA 289

RESULT 39
US-10-104-047-1637
; Sequence 1637, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1637
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1637

Query Match      68.0%; Score 17; DB 3; Length 2114;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 1172 CAAATTAAAGAAATATTGTGTAA 1196

RESULT 40
US-09-248-796A-740
; Sequence 740, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 740
; LENGTH: 2811
```

```
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-740

Query Match      68.0%; Score 17; DB 3; Length 2811;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 190 CAAATTATAGTGACTACTCTCTGAAA 214

RESULT 41
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-94

Query Match      68.0%; Score 17; DB 3; Length 8195;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 1272 CAAATCAAGGAAACTATCGTGCGA 1296

RESULT 42
US-09-949-016-12015
; Sequence 12015, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12015  
; LENGTH: 44064  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12015

Query Match 68.0%; Score 17; DB 3; Length 44064;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 26935 CAAATTATGAGCAATTATGCAA 26959

## RESULT 43

US-09-949-016-15757  
; Sequence 15757, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15757  
; LENGTH: 44072  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15757

Query Match 68.0%; Score 17; DB 3; Length 44072;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 26935 CAAATTATGAGCAATTATGCAA 26959

## RESULT 44

US-09-949-016-13502  
; Sequence 13502, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13502  
; LENGTH: 88758  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13502

Query Match 68.0%; Score 17; DB 3; Length 88758;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 6218 CACCTTAAACAGACTTTTCGTGAAA 6242

## RESULT 45

US-09-949-016-13658/c  
; Sequence 13658, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13658  
; LENGTH: 96845  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13658

Query Match 68.0%; Score 17; DB 3; Length 96845;  
Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 36152 CAAATTAAAGACACTTTTATGTAA 36128

## RESULT 46

US-09-949-016-15725/c  
; Sequence 15725, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15725  
; LENGTH: 264358  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(264358)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15725

Query Match 68.0%; Score 17; DB 3; Length 264358;  
Best Local Similarity 80.0%; Pred. No. 3.8e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
Db 167433 CACATTAAAGAACTATTGAGCAA 167409

## RESULT 47

US-09-325-932A-133  
; Sequence 133, Application US/09325932A  
; Patent No. 6451604  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; APPLICANT: Lasham, Annette  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop  
; FILE REFERENCE: 1022  
; CURRENT APPLICATION NUMBER: US/09/325,932A  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 133  
; LENGTH: 346  
; TYPE: DNA  
; ORGANISM: Pinus radiata  
US-09-325-932A-133

Query Match 67.2%; Score 16.8; DB 3; Length 346;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTG 20  
Db 89 CAAATCAAGAGGCTATTCTG 108

## RESULT 48

US-09-325-932A-132  
; Sequence 132, Application US/09325932A  
; Patent No. 6451604  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; APPLICANT: Lasham, Annette  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop  
; FILE REFERENCE: 1022  
; CURRENT APPLICATION NUMBER: US/09/325,932A  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 132  
; LENGTH: 1653  
; TYPE: DNA  
; ORGANISM: Pinus radiata  
US-09-325-932A-132

Query Match 67.2%; Score 16.8; DB 3; Length 1653;  
Best Local Similarity 90.0%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTG 20

Db 335 CAAATCAAGAGGCTATTCTG 354

## RESULT 49

US-09-798-635A-10  
; Sequence 10, Application US/09798635A  
; Patent No. 6936432  
; GENERAL INFORMATION:  
; APPLICANT: Gopalan, Venkat  
; APPLICANT: Jovanovic, Milan  
; APPLICANT: Eder, Paul S.  
; APPLICANT: Giordano, Tony  
; APPLICANT: Powers, Gordon D.  
; APPLICANT: Xavier, K. Ashish  
; TITLE OF INVENTION: Their Use in Identifying Antibacterial Compounds  
; FILE REFERENCE: 50093/016002  
; CURRENT APPLICATION NUMBER: US/09/798,635A  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 09/516,061  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 462  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-798-635A-10

Query Match 66.4%; Score 16.6; DB 3; Length 462;  
Best Local Similarity 82.6%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTCTGCGAA 25  
Db 36 AATTAAGGGGACTATTCTTGCA 58

## RESULT 50

US-09-621-976-18917  
; Sequence 18917, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 18917  
; LENGTH: 506  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-18917

Query Match 66.4%; Score 16.6; DB 3; Length 506;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCTGCGA 24  
Db 131 AATTCGAAAGACTATTCTGCA 153

## RESULT 51

US-09-270-767-4638/c  
; Sequence 4638, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:

```
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4638
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-4638

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTCTGCA 24
      ||||| ||| ||| ||| |||
Db      276 AAATTAACAACAATCTTGCA 254
      ||||| ||| ||| ||| |||

RESULT 52
US-09-270-767-19920/c
; Sequence 19920, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19920
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-19920

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTCTGCA 24
      ||||| ||| ||| ||| |||
Db      276 AAATTAACAACAATCTTGCA 254
      ||||| ||| ||| ||| |||

RESULT 53
US-09-949-016-34577
; Sequence 34577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34577
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-34577

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTCTGCA 24
      ||||| ||| ||| ||| |||
Db      276 AAATTAACAACAATCTTGCA 254
      ||||| ||| ||| ||| |||

RESULT 55
US-09-949-016-48887/c
; Sequence 48887, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48887
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48887

Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTCTGCA 24
      ||||| ||| ||| ||| |||
Db      177 AAATTAAGATAGTATTATGCA 199
      ||||| ||| ||| ||| |||

RESULT 54
US-09-949-016-34578
; Sequence 34578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34578
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-34578

Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTCTGCA 24
      ||||| ||| ||| ||| |||
Db      376 AAATTAAGATAGTATTATGCA 398
      ||||| ||| ||| ||| |||

RESULT 55
US-09-949-016-48887/c
; Sequence 48887, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48887
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48887

Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
|||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 56

US-09-949-016-48888/c  
; Sequence 48888, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48888  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48888

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
|||||  
Db 278 CAAATTACAGAGACTGCCCGTGC 256

## RESULT 57

US-09-949-016-49059/c  
; Sequence 49059, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49059  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49059

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
|||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 58

US-09-949-016-49060/c  
; Sequence 49060, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49060  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49060

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
|||||  
Db 278 CAAATTACAGAGACTGCCCGTGC 256

## RESULT 59

US-09-949-016-49231/c  
; Sequence 49231, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49231  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49231

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
|||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 60

US-09-949-016-49232/c  
; Sequence 49232, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49232  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49232

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTGCTGC 23  
Db 278 CAAATTACAGAGACTGCCCGTGC 256

Search completed: January 28, 2006, 01:18:44  
Job time : 77.2247 secs

This Page Blank (uspto)



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 433.708 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaataaagagactattctgca 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/us07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/us08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/us09\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/us09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/us10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/us10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/us10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/us10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/us10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/us11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	9	US-10-716-005-3
2	20.2	80.8	525	7	US-10-282-122A-34908
3	20.2	80.8	1731	7	US-10-282-122A-36131
4	18.8	75.2	1417	7	US-10-425-114-20074
C 5	18.6	74.4	651	5	US-10-027-632-201516
C 6	18.6	74.4	651	6	US-10-027-632-201516
C 7	18.6	74.4	168153	9	US-10-756-149-1933
C 8	18.2	72.8	1482	4	US-09-925-065A-551673
9	17.8	71.2	1344	6	US-10-369-493-40385
10	17.8	71.2	1344	7	US-10-282-122A-10106
C 11	17.6	70.4	25	7	US-10-719-956-210524
C 12	17.6	70.4	269	8	US-10-425-115-34598
C 13	17.6	70.4	271	7	US-10-437-963-5617
C 14	17.6	70.4	393	4	US-09-925-065A-515562
15	17.6	70.4	424	7	US-10-242-535A-39738
16	17.6	70.4	424	7	US-10-085-783A-39738
C 17	17.6	70.4	427	3	US-09-991-936-1224
C 18	17.6	70.4	427	9	US-10-978-245-1224
19	17.6	70.4	489	7	US-10-644-659A-7
20	17.6	70.4	514	4	US-09-925-065A-638093
C 21	17.6	70.4	517	4	US-09-925-065A-319988
C 22	17.6	70.4	537	4	US-09-925-065A-394501
C 23	17.6	70.4	586	4	US-09-925-065A-319987

ALIGNMENTS

RESULT 1  
US-10-716-005-3  
; Sequence 3, Application US/10716005  
; Publication NO. US20050106578A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; APPLICANT: Cockerill III, Franklin R.  
; APPLICANT: Aichinger, Christian  
; APPLICANT: Reiser, Astrid  
; TITLE OF INVENTION: Detection of Group B Streptococcus  
; FILE REFERENCE: 07039/460001  
; CURRENT APPLICATION NUMBER: US/10/716,005  
; CURRENT FILING DATE: 2003-11-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-716-005-3

Query Match 100.0%; Score 25; DB 9; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGCA 25

Db 1 CAAATTAAGAGACTATTCTGCA 25

Sequence 203129, A  
Sequence 61594, A  
Sequence 61595, A  
Sequence 891228, A  
Sequence 865985, A  
Sequence 8199, Ap  
Sequence 21294, A  
Sequence 3797, Ap  
Sequence 3484, Ap  
Sequence 57374, A  
Sequence 4939, Ap  
Sequence 34396, A  
Sequence 326, App  
Sequence 14998, A  
Sequence 355, App  
Sequence 1282, Ap  
Sequence 83, Appl  
Sequence 122, App  
Sequence 130, App  
Sequence 10, Appl  
Sequence 9592, Ap  
Sequence 58172, A  
Sequence 114621, A  
Sequence 114622, A  
Sequence 114623, A  
Sequence 114622, A  
Sequence 114623, A  
Sequence 35, Appl  
Sequence 18587, A  
Sequence 507, App  
Sequence 11, Appl  
Sequence 2308, Ap  
Sequence 9748, Ap  
Sequence 2308, Ap  
Sequence 1810, Ap  
Sequence 988, Ap

4

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 4 ATTAAGAGACTATTCTGTCAA 25  
Db 926 ATTAAGAAACTATTCTGTAA 947

RESULT 5  
US-10-027-632-201516/c  
; Sequence 201516, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 201516  
; LENGTH: 651  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-201516

Query Match 74.4%; Score 18.6; DB 5; Length 651;  
Best Local Similarity 84.0%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 544 CAAATTAAGAGATTATTTAGCAA 520

RESULT 6  
US-10-027-632-201516/c  
; Sequence 201516, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002

Query Match 74.4%; Score 18.6; DB 5; Length 651;  
Best Local Similarity 84.0%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 201516  
; LENGTH: 651  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-201516

Query Match 74.4%; Score 18.6; DB 6; Length 651;  
Best Local Similarity 84.0%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 544 CAAATTAAGAGATTATTTAGCAA 520

RESULT 7  
US-10-756-149-1933/c  
; Sequence 1933, Application US/10756149  
; Publication No. US20050181375A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER  
; FILE REFERENCE: file  
; CURRENT APPLICATION NUMBER: US/10/756,149  
; CURRENT FILING DATE: 2004-01-12  
; NUMBER OF SEQ ID NOS: 5818  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1933  
; LENGTH: 168153  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-756-149-1933

Query Match 74.4%; Score 18.6; DB 9; Length 168153;  
Best Local Similarity 84.0%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 160896 CAAATTAAGAGATTATTTAGCAA 160872

RESULT 8  
US-09-925-065A-551673/c  
; Sequence 551673, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551673  
; LENGTH: 1482  
; TYPE: DNA

Query Match 74.4%; Score 18.6; DB 9; Length 168153;  
Best Local Similarity 84.0%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 160896 CAAATTAAGAGATTATTTAGCAA 160872

RESULT 8  
US-09-925-065A-551673/c  
; Sequence 551673, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551673  
; LENGTH: 1482  
; TYPE: DNA

Query Match 74.4%; Score 18.6; DB 9; Length 168153;  
Best Local Similarity 84.0%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 160896 CAAATTAAGAGATTATTTAGCAA 160872

RESULT 8  
US-09-925-065A-551673/c  
; Sequence 551673, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551673  
; LENGTH: 1482  
; TYPE: DNA

Query Match 74.4%; Score 18.6; DB 9; Length 168153;  
Best Local Similarity 84.0%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
Db 160896 CAAATTAAGAGATTATTTAGCAA 160872

RESULT 8  
US-09-925-065A-551673/c  
; Sequence 551673, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551673  
; LENGTH: 1482  
; TYPE: DNA

```

; ORGANISM: Homo sapiens
US-09-925-065A-551673

Query Match      72.8%; Score 18.2; DB 4; Length 1482;
Best Local Similarity 87.0%; Pred. No. 8.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAAGAGACTATTTCGTGC 23
||||| ||||| ||||| ||||| |||||
Db 153 CAATGAAGAGGAGCTATTTCATGC 131

RESULT 9
US-10-369-493-40385
; Sequence 40385, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40385
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-10-369-493-40385

Query Match      71.2%; Score 17.8; DB 6; Length 1344;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| ||||| |||||
Db 1166 AAATTAAGAAATTTTCGTG 1186

RESULT 10
US-10-282-122A-10106
; Sequence 10106, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forgyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335

```

```

; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10106
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Bacillus anthracis
US-10-282-122A-10106

Query Match      71.2%; Score 17.8; DB 7; Length 1344;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| ||||| |||||
Db 1166 AAATTAAGAAATTTTCGTG 1186

RESULT 11
US-10-719-956-210524/c
; Sequence 210524, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,836
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 699466
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 210524
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-719-956-210524

Query Match      70.4%; Score 17.6; DB 7; Length 25;
Best Local Similarity 83.3%; Pred. No. 9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTCAA 25
||||| ||||| ||||| ||||| |||||
Db 25 AAATTAAGAGGAGGATTCGAGCAA 2

RESULT 12
US-10-425-115-34598
; Sequence 34598, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115

```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 34598
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(269)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRF4577_131556C.1
US-10-425-115-34598
```

```
Query Match 70.4%; Score 17.6; DB 8; Length 269;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 161 AACTTAAGAGAGACTATTAATCTTCAA 184
```

```
RESULT 13
US-10-437-963-5617
; Sequence 5617, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 5617
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(271)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRF4530_12387C.1
US-10-437-963-5617
```

```
Query Match 70.4%; Score 17.6; DB 7; Length 271;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 46 AATTAAGTTAATATTCGTGCAA 69
```

```
RESULT 14
US-09-925-065A-515562
; Sequence 515562, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
```

```
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 515562
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-515562
```

```
Query Match 70.4%; Score 17.6; DB 4; Length 393;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 177 AAATTAAGAGAGACTATTAACAGCAA 200
```

```
RESULT 15
US-10-242-535A-39738
; Sequence 39738, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C. C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39738
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (52)..(52)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)..(72)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (176)..(176)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (402)..(402)
; OTHER INFORMATION: n is a, c, g, or t
```

US-10-242-535A-39738

Query Match 70.4%; Score 17.6; DB 7; Length 424;  
Best Local Similarity 80.0%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTGTCGCAA 25  
||| ||||| ||||| ||||| |||||  
DB 41 CAGATTAAATNACTATTGTCGCAA 65

RESULT 16

US-10-085-783A-39738  
; Sequence 39738, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A

; CURRENT FILING DATE: 2002-02-28

; PRIOR APPLICATION NUMBER: US 60/305,340

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/275,017

; PRIOR FILING DATE: 2001-03-12

; PRIOR APPLICATION NUMBER: US 60/271,955

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 58994

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 39738

; LENGTH: 424

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (22)..(22)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (52)..(52)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (72)..(72)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (176)..(176)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (402)..(402)

; OTHER INFORMATION: n is a, c, g, or t

US-10-085-783A-39738

Query Match 70.4%; Score 17.6; DB 7; Length 424;  
Best Local Similarity 80.0%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTGTCGCAA 25  
||| ||||| ||||| ||||| |||||  
DB 41 CAGATTAAATNACTATTGTCGCAA 65

RESULT 17

US-09-991-936-1224/c

; Sequence 1224, Application US/09991936

; Publication No. US20030073827A1

; GENERAL INFORMATION:

; APPLICANT: Brandt, Kevin S.

; APPLICANT: Gaines, Patrick J.

; APPLICANT: Stinchcomb, Dan T.

; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE  
; FILE REFERENCE: FC-6-C1  
; CURRENT APPLICATION NUMBER: US/09/991,936

; CURRENT FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US/09/543,668

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: 60/128,704

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 1959

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1224

; LENGTH: 427

; TYPE: DNA

; ORGANISM: Ctenocephalides felis

US-09-991-936-1224

Query Match 70.4%; Score 17.6; DB 3; Length 427;

Best Local Similarity 83.3%; Pred. No. 1.4e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTGTCGCAA 25

||||| ||||| ||||| ||||| |||||

DB 359 AAATTAAAGAACTATTTCATGTAA 336

RESULT 18

US-10-978-245-1224/c

; Sequence 1224, Application US/10978245

; Publication No. US20050239103A1

; GENERAL INFORMATION:

; APPLICANT: Brandt, Kevin S.

; APPLICANT: Gaines, Patrick J.

; APPLICANT: Stinchcomb, Dan T.

; APPLICANT: Wisniewski, Nancy

; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE

; FILE REFERENCE: FC-6-C1

; CURRENT APPLICATION NUMBER: US/10/978,245

; CURRENT FILING DATE: 2004-10-29

; PRIOR APPLICATION NUMBER: US/09/991,936

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US/09/543,668

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: 60/128,704

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 1959

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1224

; LENGTH: 427

; TYPE: DNA

; ORGANISM: Ctenocephalides felis

US-10-978-245-1224

Query Match 70.4%; Score 17.6; DB 9; Length 427;

Best Local Similarity 83.3%; Pred. No. 1.4e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTGTCGCAA 25

||||| ||||| ||||| ||||| |||||

DB 359 AAATTAAAGAACTATTTCATGTAA 336

RESULT 19

US-10-644-659A-7

; Sequence 7, Application US/10644659A

; Publication No. US20040086920A1

; GENERAL INFORMATION:

; APPLICANT: OLSON, ERIC

; APPLICANT: ARAI, AKIKO

; TITLE OF INVENTION: STARS - A MUSCLE-SPECIFIC ACTIN-BINDING PROTEIN

; FILE REFERENCE: MYOG:03705



; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 319987  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-319987

Query Match 70.4%; Score 17.6; DB 4; Length 586;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 475 AGATTAAAGAGACTTTAAAGTGCAA 452

RESULT 24  
US-09-925-065A-203129  
; Sequence 203129, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 203129  
; LENGTH: 589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-203129

Query Match 70.4%; Score 17.6; DB 4; Length 589;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 560 AAATTAAGAGATGATACCTGCAA 583

RESULT 25  
US-10-079-079-61594  
; Sequence 61594, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome

; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61594  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894322421\_1  
US-10-972-079-61594

Query Match 70.4%; Score 17.6; DB 9; Length 600;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 266 AAATTAAGAGACTATTAGTGAAA 289

RESULT 26  
US-10-972-079-61595  
; Sequence 61595, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; LIVESTOCK  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61595  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken 19866894322421\_2  
US-10-972-079-61595

Query Match 70.4%; Score 17.6; DB 9; Length 600;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 191 AAATTAAGAGACTATTAGTGAAA 214

RESULT 27  
US-09-925-065A-891228/c  
; Sequence 891228, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome



RESULT 29  
US-10-767-701-8199  
; Sequence 8199, Application US/10767701  
; Publication No. US2004017268A1  
; GENERAL INFORMATION:

APPLICANT: KOVALIC, David K.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
FILE REFERENCE: 38-21(53377)B  
CURRENT APPLICATION NUMBER: US/10/739,930  
CURRENT FILING DATE: 2003-12-18

```
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 3797
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-CLUSTER73803_6
US-10-739-930-3797

Query Match      70.4%; Score 17.6; DB 8; Length 1431;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAATTAAGAGACTATTTCGTCAA 25
Db      396 AAAATTAAGAGAACTACTCTTGAAA 373

RESULT 32
US-10-425-114-3484/c
; Sequence 3484, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3484
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700266667_FLI
US-10-425-114-3484

Query Match      70.4%; Score 17.6; DB 7; Length 2726;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CAAATTAAGAGACTATTTCGTGCA 24
Db      528 CAAATCAGAGAGACTTTCCGTGCA 505

RESULT 33
US-10-425-115-57374/c
; Sequence 57374, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 57374
; LENGTH: 2828
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_152319C.1
```

---

```
US-10-425-115-57374

Query Match      70.4%; Score 17.6; DB 8; Length 2828;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CAAATTAAGAGACTATTTCGTGCA 24
Db      528 CAAATCAGAGAGACTTTCCGTGCA 505

RESULT 34
US-11-097-143-4939
; Sequence 4939, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4939
; LENGTH: 10892
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-4939

Query Match      70.4%; Score 17.6; DB 10; Length 10892;
Best Local Similarity 83.3%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTTCGTCAA 25
Db      7456 AAATAATCAGACTATTTCGTCAA 7479

RESULT 35
US-10-282-122A-34396
; Sequence 34396, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
```

APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34396  
; LENGTH: 19922  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-34396

Query Match 70.4%; Score 17.6; DB 7; Length 19922;  
Best Local Similarity 83.3%; Pred. No. 2.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCGCA 24  
||| ||||| ||| ||| |||  
Db 1327 CAAATTAAGAGAGACTATTCTGCGCA 1350

RESULT 36  
US-10-724-972A-326  
; Sequence 326, Application US/10724972A  
; Publication No. US20040147734A1  
; GENERAL INFORMATION:  
; APPLICANT: Doucette-Stamm, Lynn  
; APPLICANT: Bush, David  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: PATH03-16  
; CURRENT APPLICATION NUMBER: US/10/724,972A  
; CURRENT FILING DATE: 2003-12-01  
; PRIOR APPLICATION NUMBER: 09/450,969  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 7544  
; SEQ ID NO 326  
; LENGTH: 30612  
; TYPE: DNA  
; ORGANISM: S.epidermidis  
US-10-724-972A-326

Query Match 70.4%; Score 17.6; DB 7; Length 30612;  
Best Local Similarity 83.3%; Pred. No. 2.5e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTCTGCGCA 24  
||| ||||| ||| ||| |||  
Db 12031 CAAATTAAGAGAGACTATTCTGCGCA 12054

RESULT 37  
US-11-097-143-14998  
; Sequence 14998, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14998  
; LENGTH: 49380  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-14998

Query Match 70.4%; Score 17.6; DB 10; Length 49380;  
Best Local Similarity 83.3%; Pred. No. 2.7e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTCTGCGCA 24  
||| ||||| ||| ||| |||  
Db 7624 CAAATTAAGAGAGACTATTCTGCGCA 7647

RESULT 38  
US-10-085-117-355  
; Sequence 355, Application US/10085117  
; Publication No. US20030232334A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: 529452000121  
; CURRENT APPLICATION NUMBER: US/10/085,117  
; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 361  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 355  
; LENGTH: 194945  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: variation

```

; LOCATION: (1)...(194945)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-355

```

Query Match 70.4%; Score 17.6; DB 6; Length 194945;  
Best Local Similarity 83.3%; Pred. NO. 3.2e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0;

Qy 2 AAATTAAAGAGACTATTTCGTGCAA 25  
|||||  
db 192384 AAATTAAACGGACTATTTATGCAA 19

```

RESULT 39
US-10-087-192-1282/c
; Sequence 1282, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1282
; LENGTH: 220224
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(220224)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1282

```

Query Match 70.4%; Score 17.6; DB 5; Length 220224;  
Best Local Similarity 83.3%; Pred. No. 3.2e+03;  
Matches 20: Conservative 0; Mismatches 4; Indels 0;

QY 2 AAATTAAAGAGACTATTTCGTGCA 25  
|||||  
pB 150331 AAATTAAAGAGAGGATTCGGTTC 15  
|||||

```

RESULT 40
US-10-367-094-83
; Sequence 83, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandros
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367, 094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 277616
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-367-094-83

```

Query Match 70.4%; Score 17.6; DB 7; Length 277616;  
Best Local Similarity 83.3%; Pred. No. 3.3e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0;

Qy 2 AAATTAAAGAGACTATTCTGTGCA 25  
160286 AACTTAAAGAGAATATGTGTGCA 160309

```

RESULT 41
US-10-321-039-122
; Sequence 122, Application US/10321039
; Publication No. US20040014067A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Lukowiak, Andrew
; APPLICANT: Jarvis, Nancy
; APPLICANT: Kurensky, David
; TITLE OF INVENTION: Amplification Method
; FILE REFERENCE: FORS-06960
; CURRENT APPLICATION NUMBER: US/10/321039
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/998,157
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/329,113
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/360,499
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 759
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 122
; LENGTH: 1021
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (551)..(551)
; OTHER INFORMATION: n can be c o r t.
US-10-321-039-122

```

Query Match	69.6%	Score 17.4;	DB 7;	Length 1021;
Best Local Similarity	94.7%;	Pred. No. 1.9e+03;		
Matches 18:	Conservative	0;	Mismatches 1;	Indels 0;
	Gaps	0;		

QY 2 AAATTAAAGAGACTATTTCG 20  
|||  
382 AAATTAAAGAGACTATTTCG 400  
pb

```

RESULT 42
US-10-052-482-130
; Sequence 130, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 24888
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-130

```

Query Match 69.6%; Score 17.4; DB 7; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0

```

; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9592
Query Match 68.8%; Score 17.2; DB 3; Length 114;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCGTG 22
Db 27 CAAATTAAGAGACTATTTCATG 6

RESULT 45
US-10-972-079-58172/c
; Sequence 58172, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58172
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894314824_1
US-10-972-079-58172

Query Match 68.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24
Db 525 AATTAAGAGACTATACATTCA 504

RESULT 46
US-10-027-632-114621/c
; Sequence 114621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

```

; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9592
Query Match 68.8%; Score 17.2; DB 3; Length 114;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCGTG 22
Db 27 CAAATTAAGAGACTATTTCATG 6

RESULT 45
US-10-972-079-58172/c
; Sequence 58172, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58172
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894314824_1
US-10-972-079-58172

Query Match 68.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24
Db 525 AATTAAGAGACTATACATTCA 504

RESULT 46
US-10-027-632-114621/c
; Sequence 114621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

```

; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9592
Query Match 68.8%; Score 17.2; DB 3; Length 114;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCGTG 22
Db 27 CAAATTAAGAGACTATTTCATG 6

RESULT 45
US-10-972-079-58172/c
; Sequence 58172, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58172
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894314824_1
US-10-972-079-58172

Query Match 68.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24
Db 525 AATTAAGAGACTATACATTCA 504

RESULT 46
US-10-027-632-114621/c
; Sequence 114621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114621
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114621

Query Match          68.8%; Score 17.2; DB 5; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||| ||||| ||||| |||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 47
US-10-027-632-114622/c
; Sequence 114622, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114622
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114622

Query Match          68.8%; Score 17.2; DB 5; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||| ||||| ||||| |||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 48
US-10-027-632-114623/c
; Sequence 114623, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114623
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114623

Query Match          68.8%; Score 17.2; DB 6; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||| ||||| ||||| |||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 50
US-10-027-632-114622/c
; Sequence 114622, Application US/10027632
; Publication No. US20030204075A9
```

Query Match	68.8%;	Score 17.2;	DB 6;	Length 676;
Best Local Similarity	79.2%;	Pred. No. 2.2e+03;		

RESULT 53  
 US-10-437-963-18587/c  
 ; Sequence 18587, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 18587  
 ; LENGTH: 3387  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_24130C.1  
 ; US-10-437-963-18587

Query Match 68.8%; Score 17.2; DB 7; Length 3387;  
Best Local Similarity 86.4%; Pred. No. 2.7e+03;

```
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTG 22
    ||||| ||||| ||||| |||||
Db 1998 CAAATCAGAGACTATTCGTG 1977

RESULT 54
US-10-238-075-507/c
; Sequence 507, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 507
; LENGTH: 12264
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-507

Query Match 68.8%; Score 17.2; DB 6; Length 12264;
Best Local Similarity 86.4%; Pred. No. 3.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGC 23
    ||||| ||||| ||||| |||||
Db 7425 AAATTAAGTACTATTCCTGC 7404

RESULT 55
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
US-10-085-959-11

Query Match 68.8%; Score 17.2; DB 6; Length 23654;
Best Local Similarity 86.4%; Pred. No. 3.6e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGC 23
    ||||| ||||| ||||| |||||
Db 4839 AAATTAAGTACTATTCCTGC 4860
```

```
RESULT 56
US-09-764-877-2308/c
; Sequence 2308, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2308
; LENGTH: 32132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2308

Query Match 68.8%; Score 17.2; DB 3; Length 32132;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25
    ||||| ||||| ||||| |||||
Db 26595 ATTAAGAGATAATTCATGCAA 26574

RESULT 57
US-09-764-891-9748/c
; Sequence 9748, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9748
; LENGTH: 32132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9748

Query Match 68.8%; Score 17.2; DB 3; Length 32132;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25
    ||||| ||||| ||||| |||||
Db 26595 ATTAAGAGATAATTCATGCAA 26574

RESULT 58
US-10-242-515-2308/c
; Sequence 2308, Application US/10242515
; Publication No. US20040009488A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
```



Mon Jan 30 07:42:53 2006

;  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2308  
; LENGTH: 32132  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-242-515-2308

Query Match 68.8%; Score 17.2; DB 6; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 3.7e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 26595 ATTAAGAGATAATTCATGCAA 26574

RESULT 59

US-10-087-192-1810  
; Sequence 1810, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1810  
; LENGTH: 133632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-1810

Query Match 68.8%; Score 17.2; DB 5; Length 133632;  
Best Local Similarity 86.4%; Pred. No. 4.5e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24  
|||||  
Db 22221 AATAAGGAGACTATTTCATGCA 22242

RESULT 60

US-10-087-192-988/c  
; Sequence 988, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.

;  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 988  
; LENGTH: 203264  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-988  
  
Query Match 68.8%; Score 17.2; DB 5; Length 203264;  
Best Local Similarity 86.4%; Pred. No. 4.6e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 98028 ATTAAGAAACCATTAGTGCAA 98007  
  
Search completed: January 28, 2006, 01:45:40  
Job time : 438.708 secs

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 251.966 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaattaaagagactattctgcaa 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA.New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US12\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US13\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US14\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US15\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US16\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.6	70.4	2947	US-10-793-626-3805	Sequence 3805, Ap
2	17.6	70.4	3285	US-10-793-626-3153	Sequence 3153, Ap
3	17.6	70.4	3730	US-10-793-626-3499	Sequence 3499, Ap
C 4	17	68.0	2131	US-10-750-185-30811	Sequence 30811, A
C 5	17	68.0	2131	US-10-750-623-30811	Sequence 30811, A
C 6	16.6	66.4	2076	US-10-632-150-9	Sequence 9, Appli
C 7	16.6	66.4	2076	US-11-073-457-9	Sequence 9, Appli
C 8	16.6	66.4	2076	US-11-073-460-9	Sequence 9, Appli
C 9	16.6	66.4	50353	US-10-995-561-13305	Sequence 13305, A
C 10	16.6	66.4	129021	US-11-117-187-202	Sequence 202, App
11	16.6	66.4	185857	US-11-121-086-34	Sequence 34, Appl
12	16.4	65.6	201	US-10-995-561-58262	Sequence 58262, A
C 13	16.4	65.6	201	US-10-995-561-80898	Sequence 80898, A
C 14	16.4	65.6	1719	US-10-750-185-48305	Sequence 48305, A
C 15	16.4	65.6	1719	US-10-750-623-48305	Sequence 48305, A
C 16	16.4	65.6	2941	US-10-750-185-53117	Sequence 53117, A
17	16.4	65.6	2941	US-10-750-623-53117	Sequence 53117, A
C 18	16.4	65.6	21852	US-10-995-561-13452	Sequence 13452, A
C 19	16.4	65.6	43985	US-10-995-561-13337	Sequence 13337, A
C 20	16.2	64.8	201	US-10-995-561-15984	Sequence 15984, A
C 21	16.2	64.8	201	US-10-995-561-16055	Sequence 16055, A
C 22	16.2	64.8	429	US-11-136-527-1338	Sequence 1338, Ap

23	16.2	64.8	429	8	US-11-136-527-5434	Sequence 5434, Ap
24	16.2	64.8	1400	8	US-11-136-527-6155	Sequence 6155, Ap
25	16.2	64.8	1656	8	US-11-136-527-2059	Sequence 2059, Ap
C 26	16.2	64.8	11970	8	US-11-121-634-1	Sequence 1, Appli
C 27	16.2	64.8	23107	7	US-10-995-561-13201	Sequence 13201, A
C 28	16.2	64.8	168656	8	US-11-112-908-59	Sequence 59, Appl
C 29	16.2	64.8	170285	8	US-11-112-908-58	Sequence 58, Appl
C 30	16.2	64.8	199130	7	US-10-995-561-13233	Sequence 13233, A
C 31	16.2	64.8	199130	7	US-10-995-561-13233	Sequence 13233, A
C 32	16	64.0	940	7	US-10-750-185-56986	Sequence 56986, A
33	16	64.0	940	7	US-10-750-623-56986	Sequence 56986, A
34	16	64.0	1553	7	US-10-750-185-31221	Sequence 31221, A
35	16	64.0	1553	7	US-10-750-623-31221	Sequence 31221, A
C 36	16	64.0	1646	7	US-10-750-185-39461	Sequence 39461, A
C 37	16	64.0	1646	7	US-10-750-623-39461	Sequence 39461, A
C 38	16	64.0	1699	7	US-10-750-185-41491	Sequence 41491, A
C 39	16	64.0	1699	7	US-10-750-623-41491	Sequence 41491, A
C 40	16	64.0	2060	7	US-10-750-185-35978	Sequence 35978, A
C 41	16	64.0	2060	7	US-10-750-623-35978	Sequence 35978, A
C 42	16	64.0	2342	7	US-10-750-185-54406	Sequence 54406, A
C 43	16	64.0	2342	7	US-10-750-623-54406	Sequence 54406, A
C 44	16	64.0	2509	7	US-10-750-185-51187	Sequence 51187, A
45	16	64.0	2509	7	US-10-750-623-51187	Sequence 51187, A
46	16	64.0	2840	6	US-10-505-263-82	Sequence 82, Appl
C 47	16	64.0	4032	8	US-11-052-554A-663	Sequence 663, App
C 48	16	64.0	48000	8	US-11-159-597-20	Sequence 20, Appl
C 49	16	64.0	48000	8	US-11-159-597-20	Sequence 20, Appl
C 50	16	64.0	83391	8	US-11-117-187-189	Sequence 189, App
C 51	16	64.0	83391	8	US-11-117-187-189	Sequence 189, App
C 52	16	64.0	100000	8	US-11-124-368A-2881	Sequence 2881, Ap
C 53	16	64.0	159497	8	US-11-112-908-61	Sequence 61, Appl
C 54	16	64.0	160226	8	US-11-121-086-29	Sequence 29, Appl
C 55	16	64.0	171427	8	US-11-112-908-60	Sequence 60, Appl
C 56	16	64.0	189252	8	US-11-121-086-54	Sequence 54, Appl
C 57	16	64.0	195235	7	US-10-995-561-13495	Sequence 13495, A
C 58	16	64.0	212716	8	US-11-121-086-95	Sequence 95, Appl
C 59	16	64.0	1080000	7	US-10-928-446A-1	Sequence 1, Appl
60	16	64.0	1080000	7	US-10-928-446A-181	Sequence 181, App

ALIGNMENTS

RESULT 1

US-10-793-626-3805  
; Sequence 3805, Application US/10793626  
; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3805

; LENGTH: 2947

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-3805

Query Match 70.4%; Score 17.6; DB 7; Length 2947;

Best Local Similarity 83.3%; Pred. No. 54;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 24

|||||

Db 1894 CAAATTAAGAGAAATAATAGTCA 1917

```
RESULT 2
US-10-793-626-3153
; Sequence 3153, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3153
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3153

Query Match      70.4%; Score 17.6; DB 7; Length 3285;
Best Local Similarity 83.3%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CAAATTAAAGAGACTATTTCGTGCA 24
      ||||| ||||| ||||| ||||| |||||
DB      1447 CAAATAAAGAGAAATAAGTGCA 1470

RESULT 3
US-10-793-626-3499
; Sequence 3499, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3499
; LENGTH: 3730
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3499

Query Match      70.4%; Score 17.6; DB 7; Length 3730;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CAAATTAAAGAGACTATTTCGTGCA 24
      ||||| ||||| ||||| ||||| |||||
DB      1892 CAAATAAAGAGAAATAAGTGCA 1915

RESULT 4
US-10-750-185-30811/c
; Sequence 30811, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
```

```
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30811
; LENGTH: 2131
; TYPE: DNA
; ORGANISM: Bovine 19866880964537
US-10-750-185-30811

Query Match      68.0%; Score 17; DB 7; Length 2131;
Best Local Similarity 80.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 CAAATTAAAGAGACTATTTCGTGCA 25
      ||||| ||||| ||||| ||||| |||||
DB      40 CAACTCAAAGAGACTCTTTTGCTA 16

RESULT 5
US-10-750-623-30811/c
; Sequence 30811, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30811
; LENGTH: 2131
; TYPE: DNA
; ORGANISM: Bovine 19866880964537
US-10-750-623-30811

Query Match      68.0%; Score 17; DB 7; Length 2131;
Best Local Similarity 80.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 CAAATTAAAGAGACTATTTCGTGCA 25
      ||||| ||||| ||||| ||||| |||||
DB      40 CAACTCAAAGAGACTCTTTTGCTA 16

RESULT 6
US-10-632-150-9/c
; Sequence 9, Application US/10632150
; Publication No. US20050251871A1
; GENERAL INFORMATION:
; APPLICANT: Chiau, D.
; APPLICANT: Pagano, M.
; APPLICANT: Latres, E.
; TITLE OF INVENTION: NOVEL UBIQUITIN LIGASES AS THERAPEUTIC TARGETS
; FILE REFERENCE: 5914-081
```

; CURRENT APPLICATION NUMBER: US/10/632,150  
; CURRENT FILING DATE: 2003-07-30  
; PRIOR APPLICATION NUMBER: US/09/385,219  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: 60/098,355  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: 60/118,568  
; PRIOR FILING DATE: 1999-02-03  
; PRIOR APPLICATION NUMBER: 60/124,449  
; PRIOR FILING DATE: 1999-03-15  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-632-150-9

Query Match 66.4%; Score 16.6; DB 7; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 7  
US-11-073-457-9/c  
; Sequence 9, Application US/11073457  
; Publication No. US20050260556A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF PROLIFERATIVE DISORDERS  
; FILE REFERENCE: 5914-090-999  
; CURRENT APPLICATION NUMBER: US/11/073,457  
; CURRENT FILING DATE: 2005-03-04  
; PRIOR APPLICATION NUMBER: 10/042,417  
; PRIOR FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 60/260,179  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-073-457-9

Query Match 66.4%; Score 16.6; DB 8; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 8  
US-11-073-460-9/c  
; Sequence 9, Application US/11073460  
; Publication No. US20050272066A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF PROLIFERATIVE DISORDERS  
; FILE REFERENCE: 5914-090-999  
; CURRENT APPLICATION NUMBER: US/11/073,460  
; CURRENT FILING DATE: 2005-03-04  
; PRIOR APPLICATION NUMBER: 10/042,417  
; PRIOR FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 60/260,179

; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-073-460-9

Query Match 66.4%; Score 16.6; DB 8; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 9  
US-10-995-561-13305/c  
; Sequence 13305, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13305  
; LENGTH: 50353  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13305

Query Match 66.4%; Score 16.6; DB 7; Length 50353;  
Best Local Similarity 82.6%; Pred. No. 2.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 16515 AATAAAGAGATTATCGGTGTA 16493

RESULT 10  
US-11-117-187-202  
; Sequence 202, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 202  
; LENGTH: 129021  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-11-117-187-202

Query Match 66.4%; Score 16.6; DB 8; Length 129021;  
Best Local Similarity 82.6%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80898
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-80898

Query Match      65.6%; Score 16.4; DB 7; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 80808 AATTAAGAGAAATTTTCGTGCA 80830

RESULT 11
US-11-121-086-34
; Sequence 34, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: NIELSEN, KIRSTEN V.
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
; LENGTH: 165857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-34

Query Match      66.4%; Score 16.6; DB 8; Length 165857;
Best Local Similarity 82.6%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 61880 AATTAAGGAGCTGTCAAGCA 61902

RESULT 12
US-10-995-561-58262
; Sequence 58262, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58262
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-58262

Query Match      65.6%; Score 16.4; DB 7; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 8 CAAATTAAGAGACTACT 25

RESULT 13
US-10-995-561-80898/c
; Sequence 80898, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
```

```
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80898
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-80898

Query Match      65.6%; Score 16.4; DB 7; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 194 CAAATTAAGAGACTACT 177

RESULT 14
US-10-750-185-48305/c
; Sequence 48305, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48305
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-48305

Query Match      65.6%; Score 16.4; DB 7; Length 1719;
Best Local Similarity 94.4%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTC 19
   ||||| ||||| ||||| ||||| |||||
Db 289 AAATTAAGAGACTATTTC 272

RESULT 15
US-10-750-623-48305/c
; Sequence 48305, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
```

```
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 48305
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Bovine 19866880887302
US-10-750-623-48305

Query Match      65.6%; Score 16.4; DB 7; Length 1719;
Best Local Similarity 94.4%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTC 19
Db      289 AAATTAAGAGACTATTC 272

RESULT 16
US-10-750-185-53117
; Sequence 53117, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53117
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Bovine 19866880648156
US-10-750-185-53117

Query Match      65.6%; Score 16.4; DB 7; Length 2941;
Best Local Similarity 94.4%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTC 19
Db      94 AAATTAAGAGACTATTC 111

RESULT 17
US-10-750-623-53117
; Sequence 53117, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53117
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Bovine 19866880648156
US-10-750-185-53117

Query Match      65.6%; Score 16.4; DB 7; Length 2941;
Best Local Similarity 94.4%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTC 19
Db      94 AAATTAAGAGACTATTC 111

RESULT 18
US-10-995-561-13492/c
; Sequence 13492, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13492
; LENGTH: 21852
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13492

Query Match      65.6%; Score 16.4; DB 7; Length 21852;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAATTAAGAGACTATT 18
Db      10295 CAAATTAAGAGACTACT 10278

RESULT 19
US-10-995-561-13337
; Sequence 13337, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13337
; LENGTH: 43985
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13337

Query Match      65.6%; Score 16.4; DB 7; Length 43985;
Best Local Similarity 94.4%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAATTAAGAGACTATT 18
Db      1616 CAAATTAAGAGACTACT 1633
```

```
RESULT 20
US-10-995-561-15984/c
; Sequence 15984, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15984
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-15984
Query Match 64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTCAA 25
||| ||||| ||||| |||||
Db 199 TTTAGAGACTATTTCGTCAA 179

RESULT 21
US-10-995-561-16055/c
; Sequence 16055, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16055
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-16055
Query Match 64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTCAA 25
||| ||||| ||||| |||||
Db 198 TTTAGAGACTATTTCGTCAA 178

RESULT 22
US-11-136-527-1338/c
; Sequence 1338, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1338
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-1338
Query Match 64.8%; Score 16.2; DB 8; Length 429;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| |||||
Db 40 AAATTAAGAGCTCTATGCATG 20

RESULT 23
US-11-136-527-5434
; Sequence 5434, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5434
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-5434
Query Match 64.8%; Score 16.2; DB 8; Length 429;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| |||||
Db 390 AAATTAAGAGCTCTATGCATG 410

RESULT 24
US-11-136-527-6155
; Sequence 6155, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6155
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6155
Query Match 64.8%; Score 16.2; DB 8; Length 1400;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGT 21
||||| ||||| ||||| |||||
Db 1282 CAAATTAAGACTATTTCCT 1302
```



```

RESULT 25
US-11-136-527-2059
; Sequence 2059, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2059
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; US-11-136-527-2059

Query Match 64.8%; Score 16.2; DB 8; Length 1656;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 1538 CAAATAAATAGACTATTCTT 1558

RESULT 26
US-11-121-634-1/c
; Sequence 1, Application US/11121634
; Publication No. US20050282198A1
; GENERAL INFORMATION:
; APPLICANT: Duff, Gordon
; TITLE OF INVENTION: Diagnostics and Therapeutics for Diseases Associated with an IL-1
; FILE REFERENCE: 24299-508CON3 CIP1
; CURRENT APPLICATION NUMBER: US/11/121,634
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: US 10/802,061
; PRIOR FILING DATE: 2004-03-15
; PRIOR APPLICATION NUMBER: US 09/845,129
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 09/345,217
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/GB98/01481
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: GB9711040.7
; PRIOR FILING DATE: 1997-05-29
; PRIOR APPLICATION NUMBER: US 10/300,011
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/386,020
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/331,681
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/567,727
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 11970
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-121-634-1

Query Match 64.8%; Score 16.2; DB 8; Length 11970;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 88036 CAAATAAAGTACTATTTCAT 88016

RESULT 27
US-10-995-561-13201/c
; Sequence 13201, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13201
; LENGTH: 23107
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-13201

Query Match 64.8%; Score 16.2; DB 7; Length 23107;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTGCAA 25
||||| ||||| ||||| ||||| |||||
DB 11488 TTTTAGAGACTATTTGTGCAA 11468

RESULT 28
US-11-112-908-59/c
; Sequence 59, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 59
; LENGTH: 168656
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-112-908-59

Query Match 64.8%; Score 16.2; DB 8; Length 168656;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 88036 CAAATAAAGTACTATTTCAT 88016

RESULT 29
US-11-112-908-58/c
; Sequence 58, Application US/11112908

```

```
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 58
; LENGTH: 170285
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-58

Query Match      64.8%; Score 16.2; DB 8; Length 170285;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGT 21
Db 88041 CAAATTAAGTCACTATTTCAT 88021

RESULT 30
US-10-995-561-13233
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      64.8%; Score 16.2; DB 7; Length 199130;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCA 24
Db 82228 ATTGAAGAGACTATTCTTTCA 82248

RESULT 31
US-10-995-561-13233/c
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      64.8%; Score 16.2; DB 7; Length 199130;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCA 24
Db 116275 ATTGAAGAGACTATTCTTTCA 116255

RESULT 32
US-10-750-185-56986
; Sequence 56986, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56986
; LENGTH: 940
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-56986

Query Match      64.0%; Score 16; DB 7; Length 940;
Best Local Similarity 79.2%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCA 24
Db 578 CATATTAAGAGATTATTGTGAA 601

RESULT 33
US-10-750-623-56986
; Sequence 56986, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

```
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56986
; LENGTH: 940
; TYPE: DNA
; ORGANISM: Bovine 19866881765208
US-10-750-623-31221

Query Match          64.0%; Score 16; DB 7; Length 940;
Best Local Similarity 79.2%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTGCTGCA 24
   ||||| ||||| ||||| |||||
Db 578 CATATTAAAGAGATTATTGTGAA 601

RESULT 34
US-10-750-185-31221
; Sequence 31221, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31221
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Bovine 19866880903462
US-10-750-185-31221

Query Match          64.0%; Score 16; DB 7; Length 1553;
Best Local Similarity 79.2%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTGCTGCA 25
   ||||| ||||| ||||| |||||
Db 289 AAATTAAGAGACTGTGGTCCAA 312

RESULT 35
US-10-750-623-31221
; Sequence 31221, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```

```
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31221
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Bovine 19866880903462
US-10-750-623-31221

Query Match          64.0%; Score 16; DB 7; Length 1553;
Best Local Similarity 79.2%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTGCTGCA 25
   ||||| ||||| ||||| |||||
Db 289 AAATTAAGAGACTGTGGTCCAA 312

RESULT 36
US-10-750-185-39461/c
; Sequence 39461, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39461
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Bovine 19866880785007
US-10-750-185-39461

Query Match          64.0%; Score 16; DB 7; Length 1646;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTGCTGCA 24
   ||||| ||||| ||||| |||||
Db 1322 CAAATTAAGAGATTATGGAGAA 1299

RESULT 37
US-10-750-623-39461/c
; Sequence 39461, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```

```
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 39461
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Bovine 19866880785007
US-10-750-623-39461

Query Match          64.0%; Score 16; DB 7; Length 1646;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1322 CAAATTAAAGAGATTATGGGAGAA 1299

RESULT 38
US-10-750-185-41491/c
; Sequence 41491, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 41491
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Bovine 19866881235275
US-10-750-185-41491

Query Match          64.0%; Score 16; DB 7; Length 1699;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1322 CAAATTAAAGAGATTATGGGAGAA 1299

RESULT 39
US-10-750-623-41491/c
; Sequence 41491, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 41491

; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 39461
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Bovine 19866880785007
US-10-750-623-39461

Query Match          64.0%; Score 16; DB 7; Length 1646;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1322 CAAATTAAAGAGATTATGGGAGAA 1299

RESULT 40
US-10-750-185-35978/c
; Sequence 35978, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 35978
; LENGTH: 2060
; TYPE: DNA
; ORGANISM: Bovine 19866880908844
US-10-750-185-35978

Query Match          64.0%; Score 16; DB 7; Length 2060;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1033 CAAATTACAGACCATTGGTCA 1010

RESULT 41
US-10-750-623-35978/c
; Sequence 35978, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 35978
; LENGTH: 2060
; TYPE: DNA
; ORGANISM: Bovine 19866880908844
```

US-10-750-623-35978

Query Match 64.0%; Score 16; DB 7; Length 2060;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24  
||||| ||||| ||||| ||||| |||||  
Db 1033 CAAATTACAGAGACCAATTGGTCA 1010

RESULT 42

US-10-750-185-54406/c  
; Sequence 54406, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 54406  
; LENGTH: 2342  
; TYPE: DNA  
; ORGANISM: Bovine 19866880508883  
US-10-750-185-54406

Query Match 64.0%; Score 16; DB 7; Length 2342;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25  
||||| ||||| ||||| ||||| |||||  
Db 1799 AATTTAAGAGCTATTTCGACAA 1776

RESULT 43

US-10-750-623-54406/c  
; Sequence 54406, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 54406  
; LENGTH: 2342  
; TYPE: DNA  
; ORGANISM: Bovine 19866880508883  
US-10-750-623-54406

Query Match 64.0%; Score 16; DB 7; Length 2342;

Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25  
||||| ||||| ||||| ||||| |||||  
Db 1799 AATTTAAGAGCTATTTCGACAA 1776

RESULT 44

US-10-750-185-51187  
; Sequence 51187, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 51187  
; LENGTH: 2509  
; TYPE: DNA  
; ORGANISM: Bovine 19866881027540  
US-10-750-185-51187

Query Match 64.0%; Score 16; DB 7; Length 2509;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25  
||||| ||||| ||||| ||||| |||||  
Db 2455 AATCAACAGACTATTTCATTGAA 2478

RESULT 45

US-10-750-623-51187  
; Sequence 51187, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 51187  
; LENGTH: 2509  
; TYPE: DNA  
; ORGANISM: Bovine 19866881027540  
US-10-750-623-51187

Query Match 64.0%; Score 16; DB 7; Length 2509;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 2455 AAATCAAGAGACTATTTCATTGAA 2478

## RESULT 46

US-10-505-263-82  
; Sequence 82, Application US/10505263  
; Publication No. US20060014940A1  
; GENERAL INFORMATION:  
; APPLICANT: Vanderbilt University  
; APPLICANT: Case Western Reserve University  
; APPLICANT: The Brigham and Women's Hospital, Inc.  
; APPLICANT: Mount, David B  
; APPLICANT: Romero, Michael  
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF SLC26A6, SLC26A1, and SLC26A2  
; FILE REFERENCE: 1242/50/2 PCT/US  
; CURRENT APPLICATION NUMBER: US/10/505,263  
; PRIOR FILING DATE: 2004-08-20  
; PRIOR APPLICATION NUMBER: US 60/360,275  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: PCT/US03/06469  
; PRIOR FILING DATE: 2003-02-28  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 82  
; LENGTH: 2640  
; TYPE: DNA  
; ORGANISM: Xenopus laevis  
US-10-505-263-82

Query Match 64.0%; Score 16; DB 6; Length 2640;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24  
|||||  
Db 1355 CAAATTAAGGAATGTTTCATGCA 1378

## RESULT 47

US-11-052-554A-663/c  
; Sequence 663, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 663  
; LENGTH: 4032  
; TYPE: DNA  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-663

Query Match 64.0%; Score 16; DB 8; Length 4032;  
Best Local Similarity 79.2%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 1637 AAATCCAGAGACTATCCGTTGAA 1614

## RESULT 48

US-11-159-597-20  
; Sequence 20, Application US/11159597  
; Publication No. US20050255559A1  
; GENERAL INFORMATION:  
; APPLICANT: Uebele, Victor N.  
; APPLICANT: Swanson, Richard J.  
; APPLICANT: Liu, Yuan  
; APPLICANT: Lagrutta, Armando  
; TITLE OF INVENTION: NOVEL HUMAN CALCIUM SENSITIVE POTASSIUM  
; FILE REFERENCE: 20499P  
; CURRENT APPLICATION NUMBER: US/11/159,597  
; CURRENT FILING DATE: 2005-06-23  
; PRIOR APPLICATION NUMBER: US/10/031,691  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: PCT/US00/19585  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/144,764  
; PRIOR FILING DATE: 1999-07-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 48000  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(48000)  
; OTHER INFORMATION: n = A,T,C or G  
US-11-159-597-20

Query Match 64.0%; Score 16; DB 8; Length 48000;  
Best Local Similarity 79.2%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24  
|||||  
Db 3740 CAAATTAAGACCAITTTGCGCA 3763

## RESULT 49

US-11-159-597-20/c  
; Sequence 20, Application US/11159597  
; Publication No. US20050255559A1  
; GENERAL INFORMATION:  
; APPLICANT: Uebele, Victor N.  
; APPLICANT: Swanson, Richard J.  
; APPLICANT: Liu, Yuan  
; APPLICANT: Lagrutta, Armando  
; TITLE OF INVENTION: NOVEL HUMAN CALCIUM SENSITIVE POTASSIUM  
; FILE REFERENCE: 20499P  
; CURRENT APPLICATION NUMBER: US/11/159,597  
; CURRENT FILING DATE: 2005-06-23  
; PRIOR APPLICATION NUMBER: US/10/031,691  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: PCT/US00/19585  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/144,764  
; PRIOR FILING DATE: 1999-07-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 48000  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(48000)  
; OTHER INFORMATION: n = A,T,C or G  
US-11-159-597-20

Query Match 64.0%; Score 16; DB 8; Length 48000;  
Best Local Similarity 79.2%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCTGCA 25  
DB 15508 ATATGAAGAGCGTATTACTGCA 15485

## RESULT 50

US-11-117-187-189/c  
; Sequence 189, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 189  
; LENGTH: 83391  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (2556)..(2605)  
; OTHER INFORMATION: N = A, C G, or T/U  
US-11-117-187-189

Query Match 64.0%; Score 16; DB 8; Length 83391;  
Best Local Similarity 79.2%; Pred. No. 4.6e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 24  
DB 5781 CAACTTGAAGAGCGATTCTGGCA 5758

## RESULT 51

US-11-117-187-195/c  
; Sequence 195, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 195  
; LENGTH: 90336  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-11-117-187-195

Query Match 64.0%; Score 16; DB 8; Length 90336;  
Best Local Similarity 79.2%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 24

DB 18537 CAACTTGAAGAGCGATTCTGGCA 18514

## RESULT 52

US-11-124-368A-2881/c  
; Sequence 2881, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: PaatSeq for Windows Version 4.0  
; SEQ ID NO 2881  
; LENGTH: 100000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2881

Query Match 64.0%; Score 16; DB 8; Length 100000;  
Best Local Similarity 79.2%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGCA 25  
DB 54468 AAATTAAGAAACTATTGATGAA 54445

## RESULT 53

US-11-112-908-61/c  
; Sequence 61, Application US/11112908  
; Publication No. US20050260859A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 61  
; LENGTH: 159497  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-61

Query Match 64.0%; Score 16; DB 8; Length 159497;  
Best Local Similarity 79.2%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGCA 25  
DB 155485 AAATAAAGAGATTATTCTTCAA 155462

```
RESULT 54
US-11-121-086-29
; Sequence 29, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 29
; LENGTH: 160226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-29

Query Match 64.0%; Score 16; DB 8; Length 160226;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCTGTGCA 25
Db 83051 AAAAAAAGAGACTATTTTGCTA 83074

RESULT 55
US-11-112-908-60/c
; Sequence 60, Application US/11112908
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
; LENGTH: 171427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-60

Query Match 64.0%; Score 16; DB 8; Length 171427;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCTGTGCA 25
Db 61621 AAATAAAGAGACTATTTCTTCA 61598

RESULT 56
US-11-121-086-54/c
; Sequence 54, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
```

```
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 54
; LENGTH: 189252
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-54

Query Match 64.0%; Score 16; DB 8; Length 189252;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCTGTGCA 25
Db 31645 AAGTAAAGGAGATTATTCATGGA 31622

RESULT 57
US-10-995-561-13495/c
; Sequence 13495, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13495
; LENGTH: 195235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13495

Query Match 64.0%; Score 16; DB 7; Length 195235;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGTGCA 24
Db 173960 CAAAGTAAAGAGACAATCTGTGAA 173937

RESULT 58
US-11-121-086-95/c
; Sequence 95, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 95
; LENGTH: 212716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-95
```



```
Query Match      64.0%; Score 16; DB 8; Length 212716;
Best Local Similarity 79.2%; Pred. No. 5.1e+02;
Matches 19: Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

Qy 2 AAATTAAAGAGACTATTCTGTCAA 25  
||| ||| ||| ||| ||| ||| ||| |||  
Dp 22855 AAAAAAAAAGACTATGCCTGCAA 22832

RESULT 59  
 US-10-928-446A-1  
 ; Sequence 1, Application US/10928446A  
 ; Publication No. US20050277123A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION  
 ; TITLE OF INVENTION: VARIANTS OF REDDAL ASSOCIATED WITH HYPERTENSION AND  
 ; TITLE OF INVENTION: VIRAL BUDDING  
 ; FILE REFERENCE: 0274-5785.IUS  
 ; CURRENT APPLICATION NUMBER: US/10/928,446A  
 ; CURRENT FILING DATE: 2004-08-26  
 ; PRIOR APPLICATION NUMBER: 60/359,741  
 ; PRIOR FILING DATE: 2002-02-26  
 ; NUMBER OF SEQ ID NOS: 202  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1080000

```

; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825234)
; OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825270)
; OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825401)
; OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825428)
; OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825473)
; OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825765)
; OTHER INFORMATION: a "c" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825828)
; OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826041)
; OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826546)
; OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826654)
; OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826826)
; OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'q'
;
; LOCATION: (871168)
; OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872678)
; OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872742)
; OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (925859)
; OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (933220)
; OTHER INFORMATION: the 'n' at position 933220 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (993254)
; OTHER INFORMATION: the 'n' at position 993254 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1006462)
; OTHER INFORMATION: the 'n' at position 1006462 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1007820)..(1007823)
; OTHER INFORMATION: "ttct" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1018038)
; OTHER INFORMATION: the 'n' at position 1018038 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1018704)
; OTHER INFORMATION: the 'n' at position 1018704 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1018718)..(1018720)
; OTHER INFORMATION: "gtt" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1026786)
; OTHER INFORMATION: the 'n' at position 1026786 may be 'c' or 'a'

```

```
; NAME/KEY: allele
; LOCATION: (1047134)
; OTHER INFORMATION: the 'n' at position 1047134 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047159)
; OTHER INFORMATION: the 'n' at position 1047159 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047378)
; OTHER INFORMATION: the 'n' at position 1047378 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047739)
; OTHER INFORMATION: the 'n' at position 1047739 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1050133)..(1050137)
; OTHER INFORMATION: "ttaa" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1050539)
; OTHER INFORMATION: the 'n' at position 1050539 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1062808)
; OTHER INFORMATION: the 'n' at position 1062808 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1066392)
; OTHER INFORMATION: the 'n' at position 1066392 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1073711)
; OTHER INFORMATION: the 'n' at position 1073711 may be 'c' or 't'
;
US-10-928-446A-1
```

```
Query Match      64.0%; Score 16; DB 7; Length 1080000;
Best Local Similarity 79.2%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      2  AAATTAAGAGACTATTCGTGCAA 25
         ||||| ||||| ||||| |||||
Db       623021  AAAATAAGGAATAATTGTGCAA 623044
```

```
RESULT 60
US-10-928-446A-181
; Sequence 181, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; FILE OF INVENTION: VIRAL BUDDING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 181
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (826985)..(827008)
; FEATURE:
; OTHER INFORMATION: full exon 1 range is 826667-827008
; FEATURE:
; NAME/KEY: allele
; LOCATION: (827008)..(827008)
```

```
;
; OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (843242)..(843315)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (922549)..(922630)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (926021)..(926059)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (929123)..(929176)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (993104)..(993154)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (999547)..(999608)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1000354)..(1000456)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1002118)..(1002284)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1006117)..(1006249)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1007860)..(1008036)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1010940)..(1011014)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018160)..(1018291)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018800)..(1018919)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020028)..(1020225)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1026659)..(1026736)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1028113)..(1028167)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1034316)..(1034374)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1041390)..(1041455)
; OTHER INFORMATION: exon
```

Mon Jan 30 07:42:53 2006

```
FEATURE:
NAME/KEY: CDS
LOCATION: (1043121)..(1043350)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1044868)..(1044989)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1047519)..(1047589)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050296)..(1050391)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1060368)..(1060441)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062648)..(1062708)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1064561)..(1064620)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1066207)..(1066314)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1067768)..(1067864)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1068609)..(1068681)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1073289)..(1073388)
OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-181
```

Query Match 64.0%; Score 16; DB 7; Length 1080000;  
Best Local Similarity 79.2%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25  
Db 623021 AAAATAAGGAGATATTGATGCAA 623044

Search completed: January 28, 2006, 02:00:47  
Job time : 259.216 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 1620.9 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-4  
Perfect score: 20  
Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues  
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_gse1:\*  
10: gb\_gse2:\*  
11: gb\_gse3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	163	5	BU581563
C 2	18.4	92.0	482	2	BF654954
C 3	18.4	92.0	582	10	CZ263299
4	18.4	92.0	591	3	BP159851
5	18.4	92.0	599	3	BI770737
6	18	90.0	484	9	AQ312469
7	17.4	87.0	183	7	CK593526
8	17.4	87.0	197	9	AZ464666
9	17.4	87.0	274	1	AL649495
C 10	17.4	87.0	336	1	AL679022
11	17.4	87.0	350	5	EX689357
12	17.4	87.0	412	2	BE666884
C 13	17.4	87.0	429	1	AW035216
C 14	17.4	87.0	429	9	BH9397
C 15	17.4	87.0	433	1	AL966115
C 16	17.4	87.0	438	9	AQ636375
C 17	17.4	87.0	472	3	BP909630
C 18	17.4	87.0	473	5	BW554910
C 19	17.4	87.0	478	2	BF409425
C 20	17.4	87.0	478	3	BP906911
C 21	17.4	87.0	480	9	AZ961721
C 22	17.4	87.0	489	1	AL782393

C 23	17.4	87.0	490	7	CO335661
C 24	17.4	87.0	491	9	AQ013410
C 25	17.4	87.0	505	1	AL672590
C 26	17.4	87.0	525	3	BP908357
C 27	17.4	87.0	526	1	AL655162
28	17.4	87.0	541	10	CE293235
29	17.4	87.0	550	3	BP131868
30	17.4	87.0	561	3	BP721011
C 31	17.4	87.0	561	9	BZ292732
C 32	17.4	87.0	562	9	AZ596869
C 33	17.4	87.0	564	10	AG292825
34	17.4	87.0	568	10	CS03452
35	17.4	87.0	571	8	CS997388
C 36	17.4	87.0	572	8	CK997387
C 37	17.4	87.0	576	7	CK721858
C 38	17.4	87.0	586	1	AL804540
C 39	17.4	87.0	586	7	CN109679
C 40	17.4	87.0	600	5	BU919936
C 41	17.4	87.0	604	1	AL644275
C 42	17.4	87.0	604	2	BG634601
C 43	17.4	87.0	607	10	CZ498169
C 44	17.4	87.0	610	5	BX764191
C 45	17.4	87.0	613	1	AL793267
C 46	17.4	87.0	618	1	AL867548
C 47	17.4	87.0	619	3	BP685102
C 48	17.4	87.0	620	1	AL661458
C 49	17.4	87.0	620	1	AL783961
C 50	17.4	87.0	621	5	BQ521386
C 51	17.4	87.0	623	10	AG931878
C 52	17.4	87.0	634	8	CK466293
C 53	17.4	87.0	638	7	CN109680
54	17.4	87.0	638	9	AZ600504
55	17.4	87.0	639	5	BX758905
C 56	17.4	87.0	649	1	AL895824
57	17.4	87.0	652	7	CR426638
C 58	17.4	87.0	658	7	CN084914
59	17.4	87.0	663	3	BJ076021
60	17.4	87.0	669	8	CK491676

ALIGNMENTS

RESULT 1	BU581563	163 bp	mRNA	linear	EST 17-SEP-2002
LOCUS	BU581563	in38c09.y1 Human Fetal Pancreas	1B Homo sapiens cDNA 5' similar to		
DEFINITION	LOCUS	SW:MITC2 HUMAN P14209 T-CELL SURFACE GLYCOPROTEIN E2 PRECURSOR ;	mRNA sequence.		
ACCESSION	BU581563	GI:23066823			
VERSION	BU581563.1	GI:23066823			
KEYWORDS	EST.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 163)				
AUTHORS	Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blaisdin, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishevili, R., Williams, J., Jackson, Y. and Bowers, Y.				
TITLE	Endocrine Pancreas Consortium				
JOURNAL	Unpublished (2000)				
COMMENT	Other ESTs: in38c09.x1				
	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue				
	Endocrine Pancreas Consortium				
	Harvard University, Howard Hughes Medical Institute				
	Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138				
	Tel: 617-495-1812				

```

Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@imgate.wustl.edu)
Trace considered overall poor quality
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..163
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20
weeks, Stratagene #738023)"
/dev_stage="Fetal Pancreas"
/clone_lib="Human Fetal Pancreas 1B"
/note="Vector: pBluescript SK(-); Site 1: NotI; Site 2:
XhoI; cDNA made by oligo-dT priming. Size-selected on
agarose gel. Average insert size ~1kb. 5' XhoI site was
destroyed after directional cloning. Amplified once.
Contact information: Hiroshi Inoue, MD, Metabolism Div.
(Alan Permutt Lab), Washington University School of
Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO
63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,
Fax:314-747-2692."

FEATURES
source
Query Match 92.0%; Score 18.4; DB 5; Length 163;
Best Local Similarity 95.0%; Pred. No. 7.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
|||||
Db 113 CAAGAAATGCAGAAACAGG 132
|||||

RESULT 2
BF654954/c
LOCUS BF654954 482 bp mRNA linear EST 25-APR-2001
DEFINITION 279320 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF654954
VERSION BF654954.1 GI:11920086
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 482)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Ferrea,G., Holt,I., Karanymcheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 74 row: E column: 11

Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..482
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 482;
Best Local Similarity 95.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
|||||
Db 289 CAAGCAATGCAGAAACAGG 270
|||||

RESULT 3
CZ263299/c
LOCUS CZ263299 582 bp DNA linear GSS 28-FEB-2005
DEFINITION OA_BBA0107F18.f OA_BBA Oryza alta genomic clone OA_BBA0107F18 5',
genomic survey sequence.
ACCESSION CZ263299
VERSION CZ263299.1 GI:60332075
KEYWORDS GSS.
SOURCE Oryza alta
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 582)
SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wisotski,M., Yost,D.,
Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
OMAP (Oryza Map Alignment Project) - Purdue University
Unpublished (2004)
Contact: Scott A. Jackson
Jackson Laboratory
Purdue University
915 W. State St., West Lafayette, IN 47907, USA
Tel: 7654963621
Fax: 7654967255
Email: sjackson@purdue.edu
Basecalling by phred version 0.020425.c. This sequence was derived
from the raw sequence read by clipping with lcy version 1.19s.
Bases 36-617 of the raw sequence (length 1085) were retained after
clipping.
Plate: 0107 row: F column: 18
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
Location/Qualifiers
1..582
/organism="Oryza alta"
/mol_type="genomic DNA"
/db_xref="taxon:52545"
/clone="OA_BBA0107F18"
/tissue_type="young leaves"
/lab_host="DH10B-T1 phage resistant"
/clone_lib="OA_BBA"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

FEATURES
source
Query Match 92.0%; Score 18.4; DB 10; Length 582;
Best Local Similarity 95.0%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

```

```

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 428 CAAGTACATGCAGAAACAGG 409

RESULT 4
LOCUS BP159851 591 bp mRNA linear EST 30-DEC-2003
DEFINITION BP159851 full-length enriched swine cDNA library, adult thymus Sus
scrofa cDNA clone THY010108A05 5', mRNA sequence.
ACCESSION BP159851
VERSION BP159851.1 GI:40409324
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.

REFERENCE 1 (bases 1 to 591)
AUTHORS Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,
Okumura, N., Hamasima, N. and Awata, T.
TITLE PEDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)
PUBMED 14681463
COMMENT Contact: Hirohide Uenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikkenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.

FEATURES
    source
        1..591
            /organism="Sus scrofa"
            /mol_type="mRNA"
            /db_xref="taxon:9823"
            /clone="THY010108A05"
            /tissue_type="thymus"
            /dev_stage="adult"
            /clone_lib="full-length enriched swine cDNA library, adult
            thymus"

ORIGIN
    Query Match 92.0%; Score 18.4; DB 3; Length 591;
    Best Local Similarity 95.0%; Pred. No. 8.4e+02;
    Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 128 CAGGTAATGCAGAAACAGG 147

RESULT 5
LOCUS BI770737 599 bp mRNA linear EST 25-SEP-2001
DEFINITION G03061189F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5210398 5',
mRNA sequence.
ACCESSION BI770737
VERSION BI770737.1 GI:15762315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 577 CAAGTAAATGCAGAAACAGG 596

RESULT 6
LOCUS AQ312469 484 bp DNA linear GSS 04-MAY-1999
DEFINITION RPC111-10214.TV RPCI-11 Homo sapiens genomic clone RPCI-11-10214,
genomic survey sequence.
ACCESSION AQ312469
VERSION AQ312469.1 GI:4044133
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE 1 (bases 1 to 484)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPC111-10214.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```

```

REFERENCE 1 (bases 1 to 599)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsapb-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1528 row: a column: 23
High quality sequence start: 5
High quality sequence stop: 563.
Location/Qualifiers
    1..599
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:5210398"
        /lab_host="DH10B"
        /clone_lib="NIH_MGC_122"
        /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
        Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
        anonymous pool of 24 week female lung, 16 week female
        spleen, and 20-22 week male spleens. Library is oligo-dT
        primed and directionally cloned (EcoRV site is destroyed
        upon cloning). Average insert size 1.4 kb, insert size
        range 1-3 kb. Library is normalized and enriched for
        full-length clones and was constructed by C. Gruber
        (Invitrogen). Research Genetics tracking code 026. Note:
        this is a NIH_MGC Library."

ORIGIN
    Query Match 92.0%; Score 18.4; DB 3; Length 599;
    Best Local Similarity 95.0%; Pred. No. 8.4e+02;
    Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 577 CAAGTAAATGCAGAAACAGG 596

RESULT 6
LOCUS AQ312469 484 bp DNA linear GSS 04-MAY-1999
DEFINITION RPC111-10214.TV RPCI-11 Homo sapiens genomic clone RPCI-11-10214,
genomic survey sequence.
ACCESSION AQ312469
VERSION AQ312469.1 GI:4044133
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE 1 (bases 1 to 484)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPC111-10214.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)

Seq primer: 17

Class: BAC ends.

#### FEATURES

source

Location/Qualifiers  
1..484  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7538979"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-10214"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPC1-11"  
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI; RPC111 Human Male BAC Library"

#### ORIGIN

Query Match 90.0%; Score 18; DB 9; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AAGTAAATGCAGAAACAG 19

Db 16 AAGTAAATGCAGAAACAG 33

#### RESULT 7

CK593526

LOCUS

DEFINITION CK593526 tcd32a01.y2 Hydra EST -Kiel 1 Hydra magnipapillata cDNA 5', mRNA linear EST 27-JAN-2004

ACCESSION CK593526

VERSION CK593526

KEYWORDS EST.

SOURCE CK593526.1 GI:41022181

ORGANISM Hydra magnipapillata

Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;

Hydridae; Hydra;

REFERENCE 1 (bases 1 to 183)

AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,

Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,

Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,

Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Teagareishvili,R.,

Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.

WashU Hydra EST Project

Unpublished (2002)

CONTACT Hans Bode

WashU Hydra EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

Library was constructed by Konstantin Khalturin, Zoologisches

Institut, Univ. Kiel, Germany Library materials provided by Thomas

Boech, Zoologisches Institut, CAU Kiel, Germany DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Hans Bode ([hzbode@uci.edu](mailto:hzbode@uci.edu))

Seq primer: -40RP from Gibco

High quality sequence stop: 183.

Location/Qualifiers

1..183

/organism="Hydra magnipapillata"

/mol\_type="mRNA"

/strain="105"

/db\_xref="taxon:6085"

/lab\_host="DH5a"

/clone\_lib="Hydra EST -Kiel 1"

/note="Vector: pSPORT1; Site\_1: Not I; Site\_2: Sal I;

#### FEATURES

source

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 183;

Best Local Similarity 94.7%; Pred. No. 2.1e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CAAGTAAATGCAGAAACAG 19

Db 142 CAAGTAAATGCAGAAACAG 160

#### RESULT 8

AZ464666

LOCUS

DEFINITION AZ464666 lmo274h09f Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0274H09 F, genomic survey sequence. linear GSS 04-OCT-2000

ACCESSION AZ464666

VERSION AZ464666.1 GI:10622791

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 197)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,

Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

Niederhauser,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0274 row: H column: 09

Seq primer: CTTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 197.

Location/Qualifiers

1..197

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0274H09"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gi|4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

#### FEATURES

source



with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 197;  
Best Local Similarity 94.7%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATCGAGAACAGG 20  
|||||  
Db 84 AAGTAAATCGAGAACAGG 102  
|||||

## RESULT 9

AL649495/c  
LOCUS  
DEFINITION  
AL649495 XGC-gastrula Xenopus tropicalis cDNA clone TGas049j03 5',  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AL649495.2 GI:38463578  
EST.

Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 274)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sahger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Dec 13, 2001 this sequence version replaced gi:17658691.

COMMENT

Contact: Huckle E

Sahger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue

Sahger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: TGas049j03.plkSP6

Sequencing primer: SP6.

## FEATURES

source

1..274

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TGas049j03"

/dev\_stage="gastrula (stages 10-5-12 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 274;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATCGAGAACAG 19

|||||

Db 21 CAAGTAAATCGAGAACAG 3

|||||

## RESULT 10

AL679022/c

LOCUS

DEFINITION

AL679022 XGC-neurula Xenopus tropicalis cDNA clone TNeu055a09 5',  
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 336)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sahger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Mar 18, 2002 this sequence version replaced gi:19535396.

COMMENT

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli DH10B

Sahger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: TNeu055a09.plcSP6

Sequencing primer: SP6.

FEATURES

source

1..336

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TNeu055a09"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 336;

Best Local Similarity 94.7%; Pred. No. 2.3e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATCGAGAACAG 19

|||||

Db 104 CAAGTAAATCGAGAACAG 86

|||||

## RESULT 11

EX689357

LOCUS

DEFINITION

EX689357 XGC-neurula Xenopus tropicalis cDNA clone TNeu055a09 3',  
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 350)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sahger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

JOURNAL

```

COMMENT
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TNeu055a09.q1kt7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from sug of poly A+ RNA from neurula.
ECORI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.
FEATURES
    source
        1..350
            /organism="Xenopus tropicalis"
            /mol_type="mRNA"
            /db_xref="taxon:8364"
            /clone="TNeu055a09"
            /dev_stage="neurula"
            /lab_host="Escherichia coli DH10B"
            /clone_lib="XGC-neurula"
            /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from neurula.
ECORI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
ORIGIN
Query Match      87.0%; Score 17.4; DB 5; Length 350;
Best Local Similarity 94.7%; Pred. No. 2.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19
|||||
Db 249 CAAGTAATGCAGAAACAG 267

RESULT 12
BE666884
LOCUS BE666884
DEFINITION BE666884 412 bp mRNA linear EST 25-APR-2001
ACCESSION BE666884
VERSION BE666884.1 GI:10027412
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Perlea,G., Holt,I., Karamyheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred 18
and -minmatch 12. Vector identified by cross_match with the -minscore 18
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCAGCAGC
Plate: 57 row: P column: 8

Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TNeu055a09.q1kt7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from sug of poly A+ RNA from neurula.
ECORI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.
FEATURES
    source
        1..412
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /clone_lib="MARC 4BOV"
            /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 412;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20
|||||
Db 79 AAGTAATGCAGAAACAGG 97

RESULT 13
AW035216/c
LOCUS AW035216
DEFINITION AW035216 429 bp mRNA linear EST 18-MAY-2001
ACCESSION AW035216
VERSION AW035216.1 GI:5893972
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Opton,J., Craven,M.B., Bowman,C.B., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
    source
        1..429
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cLEC35C9"
            /tissue_type="callus"
            /dev_stage="25-40 days old"
            /lab_host="XLI-Blue MRF"
            /clone_lib="tomato callus, TAMU"
            /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19

```

```

Db      153 CATGTAATGCAGAAACAG 135
|||||
REFERENCE
AUTHORS   B89397 429 bp DNA linear GSS 09-APR-1999
TITLE     RPCI11-24119.TPB RPCI-11 Homo sapiens genomic clone RPCI-11-24119,
JOURNAL   genomic survey sequence.
COMMENT   B89397
          B89397.1 GI:2971070
          GSS.
          Homo sapiens (human)
          ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          1 (bases 1 to 429)
          Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences for Sequence-Ready Map Building (1998)
          Unpublished (1998)
          Other_GSSs: RPCI11-24119.TV
          Contact: Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: mdadam@tigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC
          library availability, please contact Pieter de Jong
          (pieter@delong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genetics (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
          Seq primer: SP6
          Class: BAC ends.
          Location/Qualifiers
          1..429
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="GDB:7509042"
             /db_xref="taxon:9606"
             /clone="RPCI-11-24119"
             /sex="Male"
             /cell_type="Lymphocytes"
             /clone_lib="RPCI-11"
             /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
             RPCI11 Human Male BAC Library"

FEATURES
source
Query Match 87.0%; Score 17.4; DB 9; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 15
AL966115/c
LOCUS
DEFINITION AL966115 XGC-gastrula Xenopus tropicalis cDNA clone TGAs090a12 5',
mRNA sequence.
ACCESSION AL966115
VERSION AL966115.2 GI:39020253
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db      261 CAAGTAATTCAGAAACAG 243
|||||
REFERENCE
AUTHORS   AL966115 433 bp mRNA linear EST 05-DEC-2003
TITLE     AL966115 XGC-gastrula Xenopus tropicalis cDNA clone TGAs090a12 5',
JOURNAL   mRNA sequence.
COMMENT   AL966115
          AL966115.2 GI:39020253
          EST.
          Xenopus tropicalis (western clawed frog)
          SOURCE Xenopus tropicalis
          ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db      153 CATGTAATGCAGAAACAG 135
|||||
REFERENCE
AUTHORS   B89397 429 bp DNA linear GSS 09-APR-1999
TITLE     RPCI11-24119.TPB RPCI-11 Homo sapiens genomic clone RPCI-11-24119,
JOURNAL   genomic survey sequence.
COMMENT   B89397
          B89397.1 GI:2971070
          GSS.
          Homo sapiens (human)
          ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          1 (bases 1 to 429)
          Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences for Sequence-Ready Map Building (1998)
          Unpublished (1998)
          Other_GSSs: RPCI11-24119.TV
          Contact: Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: mdadam@tigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC
          library availability, please contact Pieter de Jong
          (pieter@delong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genetics (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
          Seq primer: SP6
          Class: BAC ends.
          Location/Qualifiers
          1..429
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="GDB:7509042"
             /db_xref="taxon:9606"
             /clone="RPCI-11-24119"
             /sex="Male"
             /cell_type="Lymphocytes"
             /clone_lib="RPCI-11"
             /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
             RPCI11 Human Male BAC Library"

FEATURES
source
Query Match 87.0%; Score 17.4; DB 9; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 15
AL966115/c
LOCUS
DEFINITION AL966115 XGC-gastrula Xenopus tropicalis cDNA clone TGAs090a12 5',
mRNA sequence.
ACCESSION AL966115
VERSION AL966115.2 GI:39020253
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 433)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
COMMENT   On Nov 27, 2002 this sequence version replaced gi:25789710.
          Contact: Taylor R
          Sanger Institute
          Hinxton, Cambridgeshire, CB10 1SA, UK
          Email: trop@sanger.ac.uk
          Sanger Xenopus tropicalis EST project 2001
          TROPICALIS_SEQUENCE_ID: TGAs090a12.plkasp6
          Sequencing primer: SP6
          This sequence is from a Xenopus Gene Collection (XGC) library
          constructed by Aaron M. Zorn.
          cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13
          gastrulae. EcoRI-NotI cut cDNA was then ligated into pCSI07 with
          EcoRI at the 5' end and NotI at the 3' end.
          Vector: pCSI07; Site 1: EcoRI; Site 2: NotI
          Host: Escherichia coli XL1-blue.
          Location/Qualifiers
          1..433
             /organism="Xenopus tropicalis"
             /mol_type="mRNA"
             /db_xref="taxon:8364"
             /clone="TGAs090a12"
             /dev_stage="gastrula (stages 10.5-12 mixed)"
             /lab_host="Escherichia coli XL1-blue"
             /clone_lib="XGC-gastrula"
             /note="Vector: pCSI07; Site 1: EcoRI; Site 2: NotI; cDNA
             was oligo dt primed from sug of poly A+ RNA from stages
             10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
             into pCSI07 with EcoRI at the 5' end and NotI at the 3'
             end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 433;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAATGCAGAAACAG 19
        |||||
        Db      383 CAAGTAATGCAGAAACAG 365

RESULT 16
AQ636375/c
LOCUS
DEFINITION AQ636375 438 bp DNA linear GSS 17-JUN-1999
          RPCI-11-478P12.TJ RPCI-11 Homo sapiens genomic clone
          RPCI-11-478P12, genomic survey sequence.
ACCESSION AQ636375
VERSION AQ636375.1 GI:5099010
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          1 (bases 1 to 438)
          Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
          Map Building (1997)
          Unpublished (1997)
          Contact: Shaying Zhao, William Nierman, Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: hbeetigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC

```

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)).  
Seq primer: SP6  
Class: BAC ends.

#### FEATURES

##### source

```
1. .438
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GB:7683539"
/db_xref="taxon:9606"
/clone="RPC1-11-478P12"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPC1-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
```

#### ORIGIN

```
Query Match      87.0%; Score 17.4; DB 9; Length 438;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
DB      142 AAGTAAATACAGAAACAGG 124
```

#### RESULT 17

##### LOCUS

```
BP909630/c
DEFINITION BP909630 Lycopersicon esculentum leaf Lycopersicon esculentum cDNA
clone LC16CC06 5', mRNA sequence.
```

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### COMMENT

```
BP909630      472 bp      mRNA      linear      EST 16-MAR-2005
BP909630/c    clone LC16CC06 5', mRNA sequence.

ACCESSION BP909630
VERSION BP909630.1 GI:58253101
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 472)
Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
Moez,T., Nishimura,S. and Shibata,D.
Expressed sequence tags from the laboratory-grown miniature tomato
(Lycopersicon esculentum) cultivar Micro-Tom and mining for single
nucleotide polymorphisms and insertions/deletions in tomato
cultivars
Unpublished (2005)
Contact: Daisuke Shibata
Kazusa DNA Research Institute;
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
Tel: 81-438-52-3947
Fax: 81-438-52-3948
Email: shibata@kazusa.or.jp
This clone was obtained at our laboratory.
Please visit our web site
URL:http://www.kazusa.or.jp/jmol/microtom/indexj.html (in Japanese)
URL:http://www.kazusa.or.jp/jmol/microtom/indexj.html (in English).
```

#### FEATURES

##### source

```
1. .472
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="Micro-Tom"
/db_xref="taxon:4081"
/clone="LC16CC06"
/tissue_type="leaf"
/clone_lib="Lycopersicon esculentum leaf"
```

#### ORIGIN

```
Query Match      87.0%; Score 17.4; DB 3; Length 472;
```

```
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAG 19
DB      136 CATGTAATGCAGAAACAG 118
```

#### RESULT 18

##### LOCUS

```
BW554910/c
DEFINITION BW554910 Yutaka Satou unpublished cDNA library (csga) Ciona
savignyi cDNA clone csga089j14 3', mRNA sequence.
```

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### PUBMED

##### COMMENT

##### FEATURES

##### source

##### Location/Qualifiers

##### 1. .473

##### /organism="Ciona savignyi"

##### /mol\_type="mRNA"

##### /db\_xref="taxon:51511"

##### /clones="csga089j14"

##### /dev\_stage="gastrula stage"

##### /clone\_lib="Yutaka Satou unpublished cDNA library (csga)"

##### ORIGIN

##### Query Match

##### Best Local Similarity

##### Matches

##### Conservative

##### Mismatches

##### Indels

##### Gaps

##### Length

##### DB

##### Score

##### Pred. No.

##### 2.4e+03

##### 1; Indels

##### 0; Gaps

##### 0;

##### BF409425

##### UI-R-CA1-blf-e-02-0-UI.s1

##### UI-R-CA1-blf-e-02-0-UI 3', mRNA sequence.

##### BF409425

##### BF409425.1

##### GI:11397364

##### EST.

##### SOURCE

##### ORGANISM

##### Rattus norvegicus (Norway rat)

##### Eukaryota; Chordata; Vertebrata; Euteleostomi;

##### Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

##### Sciurognathi; Muridae; Murinae; Rattus.

##### 1 (bases 1 to 478)

##### Bonaldo,M.F., Lennon,G. and Soares,M.B.

##### Normalization and subtraction: two approaches to facilitate gene

##### discovery

##### JOURNAL

##### PUBMED

##### Genome Res. 6 (9), 791-806 (1996)

##### 8889548

## COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: bento-soares@uiowa.edu  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized pons library cDNA Library Preparation: M.B. Soares Lab  
Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 21-150, >ORR1B#LTR/MaLR  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

source Location/Qualifiers  
1..478  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CA1-b1f-e-02-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-CA1"  
/note="Vector: pR7T3D-pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CA1 library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG TISSUE=pons  
TAG\_LIB=UI-R-CA1  
TAG\_SEQ=AGCAGC

## ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 478;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20

Db 157 AAGTAAATGCAGATCAGG 139

RESULT 20

BP906911/c

LOCUS BP906911 Lycopersicon esculentum leaf Lycopersicon esculentum cDNA  
DEFINITION clone LC06AA08 5', mRNA sequence.

ACCESSION BP906911

VERSION BP906911.1 GI:58250382

KEYWORDS EST.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 478)

Yamamoto, N., Tsugane, T., Watanabe, M., Yano, K., Maeda, F., Kuwata, C., Moez, T., Nishimura, S. and Shibata, D.

Expressed sequence tags from the laboratory-grown miniature tomato (Lycopersicon esculentum) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars

JOURNAL Unpublished (2005)

CONTACT: Daisuke Shibata

Kazusa DNA Research Institute;  
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
Tel: 81-438-52-3947  
Fax: 81-438-52-3948  
Email: shibata@kazusa.or.jp  
This clone was obtained at our laboratory.

Please visit our web site  
URL: <http://www.kazusa.or.jp/jcol/microtom/indexj.html> (in Japanese)  
URL: <http://www.kazusa.or.jp/jcol/microtom/indexj.html> (in English).

## FEATURES

source Location/Qualifiers

1..478

/organism="Lycopersicon esculentum"

/mol\_type="mRNA"

/cultivar="Micro-Tom"

/db\_xref="taxon:4081"

/clone="LC06AA08"

/tissue\_type="leaf"

/clone\_lib="Lycopersicon esculentum leaf"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 3; Length 478;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

Db 143 CATGTAATGCAGAAACAG 125

RESULT 21

AZ961721/c

LOCUS

DEFINITION 2M0230G13F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

2M0230G13F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

ACCESSION AZ961721

VERSION AZ961721.1 GI:13832948

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 480)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhauser, A. and Wright, D., Weies, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0230 row: G column: 13

Seq primer: CGTTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 480.

Location/Qualifiers

1..480

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0230G13"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/note="Vector: pWD42mv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 480;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19

Db 384 CAAGTAATGCAGAAACAG 366

## RESULT 22

AL782393/c

LOCUS

DEFINITION AL782393 XGC-gastrula Xenopus tropicalis cDNA clone TGas074012 5', mRNA linear EST 20-NOV-2003

ACCESSION AL782393

VERSION AL782393.2 GI:38486081

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 489)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT On Jun 25, 2002 this sequence version replaced gi:21568097.

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: [trop@sanger.ac.uk](mailto:trop@sanger.ac.uk)

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13

gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

Host: Escherichia coli XL1-blue

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas074012.plkSP6

Sequencing primer: SP6.

## FEATURES

source

1..489

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TGas074012"

/dev\_stage="gastrula (stages 10.5-12 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-gastrula"

/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA

was oligo dt primed from Sug of poly A+ RNA from stages

## ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 489;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19

Db 384 CAAGTAATGCAGAAACAG 366

## RESULT 23

CO335661/c

LOCUS

DEFINITION EN10932.5prime Exelixis FlyTag MN08 Bluescript Drosophila melanogaster cDNA clone EN10932 5, mRNA linear EST 05-OCT-2004

ACCESSION CO335661

VERSION CO335661.1 GI:49395936

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 490)

AUTHORS Nakanishi,M., Muzong,C., Peterson,E., Laufer,A., Leung,W., Platt,D.

and Swimmer,C.

EXelixis FlyTag EST Project MN08 Library

UNPUBLISHED (2004)

CONTACT: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

Plate: EN109 row: C column: 8

High quality sequence stop: 394.

## FEATURES

source

1..490

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clone="EN10932"

/cell\_line="mbn2"

/clone\_lib="Exelixis FlyTag MN08 Bluescript"

/note="Vector: pBluescript; Site\_1: NotI; Site\_2: XhoI;

oligodT primed from LPS induced mbn2 cell line."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 490;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAAACAGG 20

Db 374 AGGTAATGCAGAAACAGG 356

## RESULT 24

AQ013410/c

LOCUS

DEFINITION RPC11-24119.TKBR RPCI-11 Homo sapiens genomic clone RPCI-11-24119, genomic survey sequence.

ACCESSION AQ013410

VERSION AQ013410.1 GI:3185975

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.  
 1 (bases 1 to 491)  
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K.,  
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and  
 Venter, J.C.  
 Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
 Unpublished (1998)  
 Other GSSs: RPC111-24119.TPB RPC111-24119.TV  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPC1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
 Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tadb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tadb/humgen/bac_end_search/bac_end_search.html)  
 Class: BAC ends.  
 Location/Qualifiers  
 1..491  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7509042"  
 /db\_xref="taxon:9606"  
 /clone="RPC1-11-24119"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPC1-11"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC Library"

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 9; Length 491;  
 Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
 |||||  
 Db 242 CAAGTAAATTCAGAAACAG 224

RESULT 25  
 AL672590/c  
 LOCUS  
 DEFINITION AL672590 XGC-gastrula Xenopus tropicalis cDNA clone TGas054d02 5',  
 mRNA sequence.  
 ACCESSION AL672590  
 VERSION AL672590.2 GI:38245777  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Xenopus tropicalis (western clawed frog)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus; Silurana.  
 1 (bases 1 to 505)  
 Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.  
 Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
 Unpublished (2003)  
 On Mar 18, 2002 this sequence version replaced gi:19528946.  
 Contact: Taylor R  
 Sanger Institute  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: tropesanger.ac.uk  
 This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Aaron M. Zorn.  
 cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13  
 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
 EcoRI at the 5' end and NotI at the 3' end.  
 Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Host: Escherichia coli XL1-blue  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS SEQUENCE ID: TGas054d02.p1cSP6  
 Sequencing primer: SP6.  
 Location/Qualifiers  
 1..505  
 /organism="Xenopus tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="TGas054d02"  
 /dev\_stage="gastrula (stages 10-5-12 mixed)"  
 /lab\_host="Escherichia coli XL1-blue"  
 /clone\_lib="XGC-gastrula"  
 /note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
 was oligo dt primed from Sug of poly A+ RNA from stages  
 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
 into pCS107 with EcoRI at the 5' end and NotI at the 3'  
 end."

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 1; Length 505;  
 Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
 |||||  
 Db 408 CAAGTAAATGCAGAAACAG 390

RESULT 26  
 BP908357/c  
 LOCUS  
 DEFINITION BP908357 Lycopersicon esculentum leaf Lycopersicon esculentum cDNA  
 clone LC11DB11 5', mRNA sequence.  
 ACCESSION BP908357  
 VERSION BP908357.1 GI:58251828  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Lycopersicon esculentum (Solanum lycopersicum)  
 Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 525)  
 Yamamoto, N., Teugane, T., Watanabe, M., Yano, K., Maeda, F., Kuwata, C.,  
 Moez, T., Nishimura, S. and Shibata, D.  
 Expressed sequence tags from the laboratory-grown miniature tomato  
 (Lycopersicon esculentum) cultivar Micro-Tom and mining for single  
 nucleotide polymorphisms and insertions/deletions in tomato  
 cultivars  
 Unpublished (2005)  
 Contact: Daisuke Shibata  
 Kazusa DNA Research Institute;  
 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
 Tel: 81-438-52-3947  
 Fax: 81-438-52-3948  
 Email: shibata@kazusa.or.jp  
 This clone was obtained at our laboratory.  
 Please visit our web site  
 URL: <http://www.kazusa.or.jp/jsol/microtom/indexj.html> (in Japanese)  
 URL: <http://www.kazusa.or.jp/jsol/microtom/indexj.html> (in English).

FEATURES  
 source  
 1..525  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="Micro-Tom"  
 /db\_xref="taxon:4081"  
 /clone="LC11DB11"  
 /tissue type="leaf"  
 /clone\_lib="Lycopersicon esculentum leaf"

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 3; Length 525;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19
||| ||||| ||||| |||||
Db 129 CATGTAATGCAGAAACAG 111

RESULT 27
AL655162/c 526 bp mRNA linear EST 20-NOV-2003
LOCUS AL655162 XGC-gastrula Xenopus tropicalis cDNA clone TGas026e19 5',
DEFINITION mRNA sequence.
ACCESSION AL655162
VERSION AL655162.2 GI:38463186
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 526)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
On Dec 13, 2001 this sequence version replaced gi:17666907.
Contact: Huckle E
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGas026e19.p1kSP6
Sequencing primer: SP6.
Location/Qualifiers
1..526
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGas026e19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORGANISM

Query Match 87.0%; Score 17.4; DB 1; Length 526;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19
||| ||||| ||||| |||||
Db 315 CAAGTAATGAAGAAACAG 297

RESULT 28
CE293235 541 bp DNA linear GSS 26-SEP-2003
LOCUS CE293235 tigr-gss-dog-17000333878560 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE293235
VERSION CE293235.1 GI:36075319
KEYWORDS GSS.
SOURCE Canis familiaris (dog)

```

```

ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 541)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1..541
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 541;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAAACAG 20
||||| ||||| ||||| |||||
Db 316 AAGTAATGCAGAAACAG 334

RESULT 29
BP131868 550 bp mRNA linear EST 05-NOV-2004
LOCUS BP131868 MAT001 Nicotiana tabacum cDNA clone BY3970, mRNA sequence.
DEFINITION BP131868
ACCESSION BP131868
VERSION BP131868.1 GI:32874753
KEYWORDS EST.
SOURCE Nicotiana tabacum (common tobacco)
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 550)
Matsuoka,K., Demura,T., Galis,I., Horiguchi,T., Sasaki,M.,
Tashiro,G. and Fukuda,H.
A Comprehensive Gene Expression Analysis Toward the Understanding
of Growth and Differentiation of Tobacco BY-2 Cells
Plant Cell Physiol. 45 (9), 1280-1289 (2004)
15509851
Contact: Ken Matsuoka
Morphogenesis Research Group
RIKEN Plant Science Center
1-7-2 Suehirocho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9575
Fax: 81-45-503-9573
Email: by2@psc.riken.go.jp, URL:http://mrg.psc.riken.go.jp/strc/
The cDNA library was constructed from mRNA isolated from lag (9 h),
log (72 h) and stationary (7 days) old BY-2 cells.
Location/Qualifiers
1..550
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Bright Yellow No.2"
/db_xref="taxon:4097"

ORGANISM

```



```

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.A.,
Bachman,S., Green,C., Wright,C.L., Campos,E.J., Benson,L.D.,
Edwards,J., Liu,L., Osogawa,K., Womack,J.E., de Jong,P.J. and
Lewin,H.A.
TITLE A cattle-human comparative map built with cattle BAC-ends and human
genome sequence
JOURNAL Genome Res. 13 (8), 1966-1972 (2003)
PUBMED 12902387
COMMENT Other GSSs: CH240_114P20.TV
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by University of Illinois at Urbana
Champaign, USA with funds provided by grant No. AG202-34480-11828
from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
Genome Sequencing Initiative)
Plate: 114 row: P column: 20
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..561
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_114P20"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull 11 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 561;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 227 AAGTAAATGCTGAACAGG 209

RESULT 32
LOCUS AZ596869 562 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0410119F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0410119 F, genomic survey sequence.
ACCESSION AZ596869
VERSION AZ596869.1 GI:11719059
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Muridae; Mus.
REFERENCE 1 (bases 1 to 562)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

/clone="BY3970"
/cell_line="BY-2"
/clone_lib="MAT001"
/note="Vector: pGEM-T easy; primer: M13 forward; mRNA
obtained from lag, log and stationary phase cells"

ORIGIN
Query Match 87.0%; Score 17.4; DB 3; Length 550;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19
|||||
Db 198 CAAGTAATGCAGACACAG 216

RESULT 30
LOCUS BP721011 561 bp mRNA linear EST 19-JUL-2004
DEFINITION BP721011 Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA
library Xenopus laevis cDNA clone XL448C06ex 3', mRNA sequence.
ACCESSION BP721011
VERSION BP721011.1 GI:46069604
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 561)
AUTHORS Osada,S., Kitayama,A., Ueno,N. and Taira,M.
TITLE Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
JOURNAL Unpublished (2004)
COMMENT Contact: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
Science and Technology Corporation, Japan
7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434
Fax: 81-03-5841-4434
Email: m.taira@biol.s.u-tokyo.ac.jp,
URL:http://www.shigen.nig.ac.jp/nbrp/xenopus/est/.

FEATURES
source Location/Qualifiers
1..561
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL448C06ex"
/tissue_type="anterior neuroectoderm"
/dev_stage="late gastrula (stage 12.5)"
/clone_lib="Osada Taira anterior neuroectoderm (ANE)
PCS105 cDNA library"

ORIGIN
Query Match 87.0%; Score 17.4; DB 3; Length 561;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19
|||||
Db 174 CAAGTAATGAAGAAACAG 192

RESULT 31
LOCUS BZ922732 561 bp DNA linear GSS 12-JUN-2003
DEFINITION BZ922732 CH240_114P20 TJ CHORI-240 Bos taurus genomic clone CH240_114P20,
genomic survey sequence.
ACCESSION BZ922732
VERSION BZ922732.1 GI:31648118
KEYWORDS GSS.
SOURCE Bos taurus (cow)

```

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
 Niederhausern,A. and Wright,D.,Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0410 row: I column: 19  
 Seq primer: CGTTGTAAACACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..562  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0410119"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PMP42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 9; Length 562;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAAACAGG 20  
 ||||| ||||| ||||| |||||  
 Db 264 AAGTAAATACAGAAACAGG 282

RESULT 33  
 AG929825/c  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster DNA, clone: DME1-014C11.r.a, genomic  
 survey sequence.  
 ACCESSION  
 AG929825  
 VERSION  
 AG929825.1 GI:58451244  
 KEYWORDS  
 GSS.  
 SOURCE  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
 1

AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,  
 Toshio,T.K. and Sakaki,Y.  
 BAC end sequences of Library DME1  
 Unpublished  
 2 (bases 1 to 564)  
 Hattori,M.  
 Direct Submission  
 Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
 (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the BAC library DME1  
 For BAC library availability, please contact Masa-Toshi Yamamoto  
 (yamamoto@kit.jp).  
 Submitted (30-11-2004) by Masahira Hattori,  
 RIKEN, Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,  
 Fax:81-45-503-9170)  
 This work was done in collaboration with Yamamoto, M-T. Drosophila  
 Genetic Resource Center  
 Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan  
 Tel: 81-75-873-2660 FAX: 81-75-861-0881  
 PRIMERS  
 Sequencing : R  
 LIBRARY  
 Vector : pKS150  
 R.Site 1 : SacI  
 R.Site 2 : SacI.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..564  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="DME1-014C11.R.fa"  
 /clone\_lib="DME1 Drosophila BAC library"  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 10; Length 564;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAAACAGG 20  
 ||||| ||||| ||||| |||||  
 Db 240 AGGTAATGCAGAAACAGG 222

RESULT 34  
 CE503452  
 LOCUS  
 DEFINITION  
 tigr-gss-dog-17000327346876 Dog Library Canis familiaris genomic,  
 genomic survey sequence.  
 ACCESSION  
 CE503452  
 VERSION  
 CE503452.1 GI:36820233  
 KEYWORDS  
 GSS.  
 SOURCE  
 Canis familiaris (dog)  
 ORGANISM  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.  
 REFERENCE  
 1 (bases 1 to 568)  
 AUTHORS  
 Kirkness,B.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop.M., Wang,W., Fraser,C.M. and  
 Venter,J.C.  
 TITLE  
 JOURNAL  
 PUBMED  
 COMMENT  
 The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 14512627  
 Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200

Mon Jan 30 07:42:55 2006

Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

# FEATURES

## Location/Qualifiers

1. .568  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

# ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 568;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAGTAAATGCAGAAACAGG 20  
|||||  
Db 407 AAGTAAATGCAGAAACAGG 425  
|||||

# RESULT 35

CX997388 571 bp mRNA linear EST 09-FEB-2005  
JGI\_CAAQ1214.fwd NIH\_XGC\_trophHtl Xenopus tropicalis cDNA clone  
IMAGE:7721993 5', mRNA sequence.

CX997388  
CX997388.1 GI:58836934

# ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

1. (bases 1 to 571)

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

Other ESTs: JGI\_CAAQ1214.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Bruce Blumberg Laboratory, University of

California, Irvine

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: CAAQ 0013 row: k column: 15

High quality sequence stop: 571.

Location/Qualifiers

1. .571

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7721993"

/tissue\_type="Heart"

/dev\_stage="Adult"

/lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"

# FEATURES

## source

## Location/Qualifiers

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

/clone\_lib="NIH\_XGC\_trophHtl"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming  
(5'-ACTAGTCGGCGCTAGGCTCGAGTGTGTTTTTTTTTTTTTTT-3') and  
Stratascript reverse transcriptase. After ligation of  
EcoRI adapters (5'-AATCGGCACGAGG-3') followed by kinaasing  
adapters and by XhoI digestion, the cDNA was size selected  
by chromatography on Sepharose CL-2B columns and fractions  
containing cDNAs larger than 1000 bp were ligated into  
EcoRI/XhoI-digested pCS107. Reference for library  
construction: Current Genomics 4, 635-644. Library  
constructed by Michelle Tabb and Bruce Blumberg (Dept of  
Developmental and Cell Biology, University of California,  
Irvine)."

# ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 571;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
Db 220 CAAGTAAATGCAGAAACAG 238  
|||||

# RESULT 36

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CX997387 572 bp mRNA linear EST 09-FEB-2005  
JGI\_CAAQ1214.rev NIH\_XGC\_trophHtl Xenopus tropicalis cDNA clone  
IMAGE:7721993 3', mRNA sequence.  
CX997387  
CX997387.1 GI:58836933  
EST.  
Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

1. (bases 1 to 572)

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

Other ESTs: JGI\_CAAQ1214.fwd

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Bruce Blumberg Laboratory, University of

California, Irvine

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix '.rev'

indicates a reverse sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: CAAQ 0013 row: k column: 15

High quality sequence stop: 572.

Location/Qualifiers

1. .572

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7721993"

```

/tissue_type="Heart"
/dev_stages="Adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_lib="NIH XGC trophHrt1"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The
library was prepared from 5 ug of poly A+ RNA by oligo-dT
priming
(5'-ACTAGTGGCGCGCTAGGCTCGAGTCTTTTCTTTTCTTTTCTTTT-3') and
Stratascript reverse transcriptase. After ligation of
EcoRI adapters (5'-AATTCGACGAGG-3') followed by kinasing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on Sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested pCS107. Reference for library
construction: Current Genomics 4, 635-644. Library
constructed by Michelle Tabb and Bruce Blumberg (Dept of
Developmental and Cell Biology, University of California,
Irvine)."

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 8; Length 572;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
DB 352 CAAGTAAATGCAGAAACAG 334

```

RESULT 37  
CK721858

```

LOCUS      CK721858      576 bp      mRNA      linear      EST 10-FEB-2004
DEFINITION tad66h06.y2 Hydra EST -Kiel 1 Hydra magnipapillata cDNA 5', mRNA
sequence.

```

```

ACCESSION   CK721858
VERSION     CK721858.1 GI:42512717
KEYWORDS    EST.
SOURCE      Hydra magnipapillata

```

## ORGANISM

```

Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;
Hydridae; Hydra.

```

```

REFERENCE   1 (bases 1 to 576)

```

## AUTHORS

```

Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,
Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Teagareishvili,R.,
Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.

```

```

WashU Hydra EST Project

```

## TITLE

## JOURNAL

## COMMENT

```

Contact: Hans Bode
WashU Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

```

```

Email: est@watson.wustl.edu

```

```

Library was constructed by Konstantin Khalturin, Zoologisches
Institut, Univ. Kiel, Germany Library materials provided by Thomas
Boech, Zoologisches Institut, CAU Kiel, Germany DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Hans Bode (hrobe@edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 576.

```

## FEATURES

## source

```

1..576
Location/Qualifiers
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="105"
/db_xref="taxon:6085"
/lab_host="DH5a"
/clone_lib="Hydra EST -Kiel 1"
/notes="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I;
pSPORT 1 Vector is ampicillin resistant, M13 reverse

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 7; Length 576;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
DB 415 CAAGTAAATGCAGAAACAG 433

```

RESULT 38  
AL804540/c

```

LOCUS      AL804540      586 bp      mRNA      linear      EST 14-NOV-2003
DEFINITION AL804540 XGC-neurula Xenopus tropicalis cDNA clone TNeu081f15 5',
mRNA sequence.

```

```

ACCESSION   AL804540
VERSION     AL804540.2 GI:38330604
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)

```

## ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

```

REFERENCE   1 (bases 1 to 586)

```

## AUTHORS

```

Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)

```

## JOURNAL

```

COMMENT     On Jun 25, 2002 this sequence version replaced gi:21590908.
Contact: Taylor R
Sanger Institute

```

```

Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk

```

```

This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.

```

```

cDNA was oligo dT primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.

```

```

Vector: pCS107; Site 1: EcoRI; Site_2: NotI

```

```

Host: Escherichia coli DH10B

```

```

Sanger Xenopus tropicalis EST project 2001

```

```

TROPICALIS_SEQUENCE ID: TNeu081f15.p1cSP6

```

```

Sequencing primer: SP6.

```

```

Location/Qualifiers
1..586
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu081f15"
/dev_stages="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

```

## FEATURES

## source

```

1..586
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu081f15"
/dev_stages="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 1; Length 586;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
DB 396 CAAGTAAATGCAGAAACAG 378

```

RESULT 39  
CN109679

```

LOCUS      CN109679      586 bp      mRNA      linear      EST 31-MAR-2004

```

DEFINITION EC2CAA34AF06.b1 Xenopus tropicalis xthr plasmid library Xenopus tropicalis cDNA clone xthr34K11 3', mRNA sequence.

ACCESSION CN109679

VERSION CN109679.1 GI:45902375

KEYWORDS EST.

ORGANISM Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 586)

AUTHORS Thuret,R., Fierro,A.C., Coen,L., Perron,M., Demeneix,B., Wegnez,M., Gyapay,G., Weissenbach,J., Wincker,P., Mazabraud,A. and Pollet,N.

TITLE Exploring the nervous system transcriptome in the model Xenopus tropicalis using EST analysis

JOURNAL Unpublished (2004)

COMMENT Contact: Pollet N  
Transgenese et Genetique des Amphibiens  
CNRS UMR 8080  
IBAC bat 447, Universite Paris Sud, Orsay, F-91405, France  
Tel: +33 169157272  
Fax: +33 169156816  
Email: Nicolas.Pollet@ibaic.u-psud.fr.

Location/Qualifiers

1..586

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="Ivory Coast"

/db\_xref="taxon:8364"

/clone="xthr34K11"

/tissue type="pool of heads and retinas from tailbud stages 25-35"

/dev\_stage="stage 25-35"

/lab\_host="E.coli DH10B"

/clone\_lib="Xenopus tropicalis xthr plasmid library"

/note="Vector: pCMVSPORT6 xthr; Site 1: SfiI; Site 2: SfiI; Xenopus tropicalis polyA+ RNA was obtained from pool of heads and retinas from tailbud stages 25-35 cDNAs were synthesized using the SMART system of CLONTECH and directionally cloned into pCMVSPORT6 xthr, a modified version of pCMVSPORT6 allowing directional cloning using asymmetric SfiI sites. For antisense RNA synthesis, use T7 promoter and for sense RNA use SP6 promoter. Library constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A. Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay, France)."

ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 586;  
Best Local Similarity 94.7%; Pred.No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||

Db 183 CAAGTAAATGCAGAAACAG 201  
|||||

RESULT 40

BU919936/c

LOCUS BU919936.1 GI:24423788

DEFINITION BU919936.1 GI:24423788

ACCESSION BU919936

VERSION BU919936.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 600)

AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.

TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis

JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)

COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

Location/Qualifiers

1..600

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/tissue type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

ORIGIN

Query Match 87.0%; Score 17.4; DB 5; Length 600;  
Best Local Similarity 94.7%; Pred.No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||

Db 49 CAAGTGAATGCAGAAACAG 31  
|||||

RESULT 41

AL644275/c

LOCUS AL644275.2 GI:38222278

DEFINITION AL644275.2 GI:38222278

ACCESSION AL644275

VERSION AL644275.2

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 604)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT On Nov 7, 2001 this sequence version replaced gi:16796400.

Contact: Huckle E  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn. Sug of poly A+ RNA from neurula. EcorI-NotI cut cDNA was then ligated into pCS107 with EcorI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site 1: EcorI; Site 2: NotI  
Host: Escherichia coli DH10B  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TNeu027117.plkSP6  
Sequencing primer: SP6.

Location/Qualifiers

1..604

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TNeu027117"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/clone\_lib="XGC-neurula"

/notes="Vector: pCS107; Site 1: EcorI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from neurula. EcorI-NotI cut cDNA was then ligated into pCS107 with EcorI at the 5' end and NotI at the 3' end. EcorI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end."

```

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 604;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAG 19
|||||
Db      499 CAAGTAAATGAAGAAACAG 481

RESULT 42
BG634601/c
LOCUS      BG634601                604 bp      mRNA      linear      EST 23-APR-2001
DEFINITION AT30859.Sprime AT Drosophila melanogaster adult testes pOTB7
            Drosophila melanogaster cDNA clone AT30859 5 similar to Cytc;
            FBan006292 'cell cycle regulator' located on: 3L 74D2-74D3;
            04/09/2001, mRNA sequence.
ACCESSION  BG634601
VERSION     BG634601.1 GI:13762138
KEYWORDS   EST.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 604)
AUTHORS     Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,
            Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,
            Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
            Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J.,
            Pacieb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,
            Lewis,S.E., Celnikler,S. and Rubin,G.M.
            BDGP/HMI AT Drosophila EST Project
            Unpublished (2000)
TITLE       Contract: Stapleton, M.
JOURNAL     BDGP
COMMENT     Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
            hit genomic AE003523: arm:3L [17391057,17674393]
            estimated-cyto:74D1-75A1: 04/09/2001
            Plate: AT.308 row: E column: 11
            High quality sequence stop: 603.
            Location/Qualifiers
                length=604
                orgname="Drosophila melanogaster"
                mol_type="mRNA"
                db_xref="taxon:7227"
                clone="AT30859"
                sex="male"
                dev_stage="0-3 day old Ore-R males"
                lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
                AT.121-AT.319: DHS-alpha Tona"
                clone_lib="AT Drosophila melanogaster adult testes pOTB7"
                note="Organ: ADULT testes; Vector: pOTB7; Site 1: EcoRI;
                Site 2: XhoI. The mRNA for the testis library was made
                from testes and seminal vesicles hand dissected from 0-3
                day old Ore-R males. RNA kindly provided by the lab of
                Margaret Fuller. Sized fractionated cDNAs were directly
                ligated into pOTB7. Plasmid cDNA library."

ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 604;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 20
|||||
Db      365 AGGTAAATGCAGAAACAG 347

RESULT 43
CZ498169/c
LOCUS      CZ498169                607 bp      DNA      linear      GSS 11-MAY-2005
DEFINITION OA_BBA0142K02.f OA_BBA Oryza alta genomic clone OA_BBA0142K02 5',
            genomic survey sequence.
ACCESSION  CZ498169
VERSION     CZ498169.1 GI:63218977
KEYWORDS   GSS.
SOURCE     Oryza alta
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 607)
AUTHORS     SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wisotaki,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
            OMAP (Oryza Map Alignment Project) - Purdue University
            Unpublished (2004)
TITLE       Contact: Scott A. Jackson
JOURNAL     Jackson Laboratory
COMMENT     Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel.: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with Lucy version 1.198.
            Bases 31-637 of the raw sequence (length 992) were retained after
            clipping.
            Plate: 0142 row: K column: 02
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
            Location/Qualifiers
                length=607
                orgname="Oryza alta"
                mol_type="genomic DNA"
                db_xref="taxon:52545"
                clone="OA_BBA0142K02"
                tissue_type="young leaves"
                lab_host="DH10B-T1 phage resistant"
                clone_lib="OA_BBA"
                note="Vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 607;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAG 19
|||||
Db      375 CAAGTAAATGCAGAAACAG 357

RESULT 44
BX764191
LOCUS      BX764191                610 bp      mRNA      linear      EST 10-DEC-2003
DEFINITION BX764191 XGC-gastrula Xenopus tropicalis cDNA clone TCas090612 3',
            mRNA sequence.
ACCESSION  BX764191
VERSION     BX764191.1 GI:39671400
KEYWORDS   EST.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1 (bases 1 to 610)
AUTHORS     Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (11_2001)
            Unpublished (2003)
TITLE       Contact: Croning MDR
JOURNAL
COMMENT

```

```

Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tgas090012.q1kT7
Sequencing primer: 17
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..610
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

FEATURES
source
1..610
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 5; Length 610;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 225 CAAGTAAATGCAGAAACAG 243

RESULT 45
AL793267/c
LOCUS AL793267 613 bp mRNA linear EST 13-NOV-2003
DEFINITION AL793267 XGC-neurula Xenopus tropicalis cDNA clone TNeu108a11 5',
mRNA sequence.
ACCESSION AL793267.2 GI:38315272
VERSION AL793267.2
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 613)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
JOURNAL
COMMENT On Jun 25, 2002 this sequence version replaced gi:21578971.

Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli DH10B
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu108a11.plcSP6
Sequencing primer: SP6.
Location/Qualifiers
1..613
/organism="Xenopus tropicalis"

FEATURES
source
1..613
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 618;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 434 CAAGTAAATGCAGAAACAG 416

RESULT 46
AL967548/c
LOCUS AL967548 618 bp mRNA linear EST 05-DEC-2003
DEFINITION AL967548 XGC-gastrula Xenopus tropicalis cDNA clone Tgas131122 5',
mRNA sequence.
ACCESSION AL967548
VERSION AL967548.2 GI:39023527
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 618)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
JOURNAL
COMMENT On Nov 27, 2002 this sequence version replaced gi:25791143.

Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tgas131122.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..618
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

FEATURES
source
1..618
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 618;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 CAAGTAATGCAGAAACAG 19
Db 404 CAAGTAATGAAGAAACAG 386

RESULT 47
BP685102/c
LOCUS BP685102 Osada Taira anterior neuroectoderm (ANE) pCS105 cDNA
DEFINITION library Xenopus laevis cDNA clone XL448c06ex 5', mRNA sequence.
ACCESSION BP685102
VERSION BP685102.1 GI:46033057
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 619)
Osada,S., Kitayama,A., Ueno,N. and Taira,M.
Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
Unpublished (2004)
Contact: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434
Fax: 81-03-5841-4434
Email: m_taira@btol.s.u-tokyo.ac.jp,
URL:http://www.shigen.nig.ac.jp/nbrp/xenopus/est/.
Location/Qualifiers
1..619
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL448c06ex"
/tissue_type="anterior neuroectoderm"
/dev_stage="late gastrula (stage 12.5)"
/clone_lib="Osada Taira anterior neuroectoderm (ANE)
pCS105 cDNA library"

FEATURES
source
ORIGIN
Query Match 87.0%; Score 17.4; DB 3; Length 619;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19
Db 396 CAAGTAATGAAGAAACAG 378

RESULT 48
AL661458/c
LOCUS AL661458 Xenopus tropicalis cDNA clone TNeu03611 5',
DEFINITION mRNA sequence.
ACCESSION AL661458
VERSION AL661458.2 GI:38228185
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 620)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
COMMENT On Dec 13, 2001 this sequence version replaced gi:17676066.
Contact: Huckle E
Sanger Institute

Qy 1 CAAGTAATGCAGAAACAG 19
Db 411 CAAGTAATGAAGAAACAG 393

RESULT 49
AL783961/c
LOCUS AL783961 XGC-gastrula Xenopus tropicalis cDNA clone TGae084j16 5',
DEFINITION mRNA sequence.
ACCESSION AL783961
VERSION AL783961.2 GI:38291460
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 620)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
COMMENT On Jun 25, 2002 this sequence version replaced gi:21569665.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
cDNA was oligo dT primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TGae084j16.plkSP6
Sequencing primer: SP6.
Location/Qualifiers
1..620
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"

FEATURES
source
ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 620;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19
Db 411 CAAGTAATGAAGAAACAG 393

RESULT 49
AL783961/c
LOCUS AL783961 XGC-gastrula Xenopus tropicalis cDNA clone TGae084j16 5',
DEFINITION mRNA sequence.
ACCESSION AL783961
VERSION AL783961.2 GI:38291460
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 620)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
COMMENT On Jun 25, 2002 this sequence version replaced gi:21569665.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
cDNA was oligo dT primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TGae084j16.plkSP6
Sequencing primer: SP6.
Location/Qualifiers
1..620
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"

FEATURES
source

```

```

Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dT primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli DH10B
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TNeu036111.plkSP6
Sequencing primer: SP6.
Location/Qualifiers
1..620
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu036111"
/dev_stage="neurula"
/lab_hosts="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 620;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19
Db 411 CAAGTAATGAAGAAACAG 393

RESULT 49
AL783961/c
LOCUS AL783961 XGC-gastrula Xenopus tropicalis cDNA clone TGae084j16 5',
DEFINITION mRNA sequence.
ACCESSION AL783961
VERSION AL783961.2 GI:38291460
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 620)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
COMMENT On Jun 25, 2002 this sequence version replaced gi:21569665.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
cDNA was oligo dT primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TGae084j16.plkSP6
Sequencing primer: SP6.
Location/Qualifiers
1..620
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"

FEATURES
source

```



```

/clone="TGas084j16"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 620;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
    |||||
Db 408 CAAGTAAATGAAGAAACAG 390

RESULT 50
BQ521386      621 bp mRNA linear EST 10-JUN-2002
LOCUS        NISC n110901.y1 NTCHE XGC Emb7 Xenopus tropicalis cDNA clone
DEFINITION   IMAGE:5336233 5', mRNA sequence.
ACCESSION    BQ521386
VERSION      BQ521386.1 GI:21380255
KEYWORDS     EST.
SOURCE       Xenopus tropicalis (western clawed frog)
ORGANISM     Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 621)
NIH-XGC http://image.llnl.gov/image/html/xenopuslib.info.shtml.
National Institute of Child Health and Human Development, National
Cancer Institute, Xenopus Gene Collection
Unpublished (2002)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: L1AM11853 row: M column: 2
Seq primer: M13RP1 reverse primer (ABI).
FEATURES
source      Location/Qualifiers
1..621
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:5336233"
/tissue_type="tailbud"
/dev_stage="embryo, stages 20-27"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHG XGC Emb7"
/notes="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Invitrogen. Note: This is a
Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match      87.0%; Score 17.4; DB 5; Length 621;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
    |||||
Db 409 CAAGTAAATGAAGAAACAG 391

```

```

RESULT 51
AG931878/c
LOCUS        AG931878
DEFINITION   Drosophila melanogaster DNA, clone: DME1-017K16.F.fa, genomic
survey sequence.
ACCESSION    AG931878
VERSION      AG931878.1 GI:58453297
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
ORGANISM     Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1
Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,
Toshio,T.K. and Sakaki,Y.
BAC end sequences of Library DME1
Unpublished
2 (bases 1 to 623)
Hattori,M.
Direct Submission
Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the BAC library DME1
For BAC library availability, please contact Masa-Toshi Yamamoto
(yamamoto@kit.jp)
Submitted (30-11-2004) by Masahira Hattori,
RIKEN, Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,
Fax:81-45-503-9170)
This work was done in collaboration with Yamamoto, M-T. Drosophila
Genetic Resource Center
Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
Tel: 81-75-873-2660 FAX: 81-75-861-0881
PRIMERS
Sequencing : F
LIBRARY
Vector      : pKS150
R.Site 1    : SacI
R.Site 2    : SacI.
FEATURES
source      Location/Qualifiers
1..623
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="DME1-017K16.F.fa"
/clone_lib="DME1 Drosophila BAC library"

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 623;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAGG 20
    |||||
Db 302 AGGTAATGCAGAAACAGG 284

RESULT 52
CX466293/c
LOCUS        CX466293
DEFINITION   JGI XZG47538.fwd NIH XGC tropGae7 Xenopus tropicalis cDNA clone
IMAGE:7562072 5', mRNA sequence.
ACCESSION    CX466293
VERSION      CX466293.2 GI:71979250
KEYWORDS     EST.
SOURCE       Xenopus tropicalis (western clawed frog)
ORGANISM     Xenopus tropicalis

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

## REFERENCE

## AUTHORS

1 (bases 1 to 634)

## TITLE

## JOURNAL

## COMMENT

Richardson, P., Lucas, S., Rokhsar, D., Detter, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

On Jan 10, 2005 this sequence version replaced gi:57352169.

Other ESTs: JGI.XZG47538.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley: <http://tropicalis.berkeley.edu/home>

CDNA Library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: XZG 0493 row: d column: 6

High quality sequence stop: 634.

## FEATURES

## source

```
1..634
    /organism="Xenopus tropicalis"
    /mol_type="mRNA"
    /db_xref="taxon:8364"
    /clone="IMAGE:7562072"
    /tissue_type="whole embryo"
    /dev_stage="Gastrula (st. 10.5-12.5)"
    /lab_host="E. coli XL1-Blue derivative, Stratagene
Electrofen-Blue"
    /clone_lib="NIH XGC tropGas7"
    /note="Vector: pCS108; Site 1: SalI; Site 2: NotI;
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dt primers
(Invitrogen SuperScript plasmid System for cDNA Synthesis
and Cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
."
```

## ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 634;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

|||||

Db 436 CAAGTAAATGAAGAAACAG 418

RESULT 53

CN109680/c

LOCUS

DEFINITION

EC2CA34AF06.g1 Xenopus tropicalis xthr plasmid library Xenopus

tropicalis cDNA clone xthr34kl1 5', mRNA sequence.

ACCESSION

CN109680

VERSION

KEYWORDS

SOURCE

ORGANISM

CN109680 638 bp mRNA linear EST 31-MAR-2004

EC2CA34AF06.g1 Xenopus tropicalis xthr plasmid library Xenopus

tropicalis cDNA clone xthr34kl1 5', mRNA sequence.

CN109680

CN109680.1 GI:45902376

EST.

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

## REFERENCE

## AUTHORS

1 (bases 1 to 638)

## TITLE

## JOURNAL

## COMMENT

Thuret, R., Fierro, A.C., Coen, L., Perron, M., Demeneix, B., Wegnez, M., Gvapaty, G., Weissenbach, J., Wincker, P., Mazabraud, A. and Pollet, N.

Exploring the nervous system transcriptome in the model Xenopus

tropicalis using EST analysis

Unpublished (2004)

Contact: Pollet N

Transgenese et Genetique des Amphibiens

CNRS UMR 8080

IBAC bat 447, Universite Paris Sud, Orsay, F-91405, France

Tel: +33 169157272

Fax: +33 169156816

Email: Nicolas.Pollet@ibaic.u-psud.fr.

Location/Qualifiers

1..638

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="Ivory Coast"

/db\_xref="taxon:8364"

/clones="xthr34kl1"

/tissue\_type="pool of heads and retinas from tailbud

stages 25-35"

/dev\_stage="stage 25-35"

/lab\_host="E.coli DH10B"

/clone\_lib="Xenopus tropicalis xthr plasmid library"

/note="Vector: pCMVSPORT6.xthr; Site\_1: SfiI; Site\_2:

SfiI; Xenopus tropicalis polyA+ RNA was obtained from pool

of heads and retinas from tailbud stages 25-35 cDNAs were

synthesized using the SMART system of CLONTECH and

directionally cloned into pCMVSPORT6.xthr, a modified

version of pCMVSPORT6 allowing directional cloning using

asymmetric SfiI sites. For antisense RNA synthesis, use T7

promoter and for sense RNA use SP6 promoter. Library

constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A.

Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay,

France)."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 638;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

|||||

Db 402 CAAGTAAATGAAGAAACAG 384

RESULT 54

AZ600504

LOCUS

DEFINITION

1M0418C20F Mouse 10kb plasmid UUC1M library Mus musculus genomic

clone UUC1M0418C20 F, genomic survey sequence.

ACCESSION

AZ600504

VERSION

AZ600504.1 GI:11722694

KEYWORDS

GSS.

SOURCE

ORGANISM

Mus musculus

(house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 638)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhauser, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0418 row: C column: 20  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 638.  
 Location/Qualifiers

## FEATURES

source

```

1. .638
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="UUGC1M0418C20"
  /sex="Male"
  /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
  /clone_lib="Mouse 10kb plasmid UUGC1M library"
  /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G14732114|GB|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 9; Length 638;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 AAGTAAATGCAGAAACAGG 20

|||||

Db 280 AAGTAAATACAGAAACAGG 298

RESULT 55

BX758905

```

LOCUS      BX758905 XGC-gastrula Xenopus tropicalis cDNA clone TGas131122 3',
DEFINITION mRNA sequence.

```

ACCESSION BX758905

VERSION BX758905.1 GI:396666113

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 639)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas131122.q1kT7

Sequencing primer: T7  
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue.

## FEATURES

source

```

1. .639
  /organism="Xenopus tropicalis"
  /mol_type="mRNA"
  /db_xref="taxon:8364"
  /clone="TGas131122"
  /dev_stage="gastrula (stages 10.5-12 mixed)"
  /lab_host="Escherichia coli XL1-blue"
  /clone_lib="XGC-gastrula"
  /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 5; Length 639;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 CAAGTAAATGCAGAAACAG 19

|||||

Db 236 CAAGTAAATGAAGAAACAG 254

RESULT 56

AL895824/c

LOCUS

DEFINITION XGC-egg Xenopus tropicalis cDNA clone TEGg100k12 5', mRNA

sequence.

ACCESSION AL895824

VERSION AL895824.2 GI:38695676

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 649)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT On Sep 16, 2002 this sequence version replaced gi:22948337.

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TEGg100k12.plkSP6

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 649)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas131122.q1kT7

```

/clone.lib="XGC-egg"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 649;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAGAAACAG 19
|||||
Db 406 CAAGTAAATGAGAAACAG 388

RESULT 57
CR426638
LOCUS
DEFINITION CR426638 XGC-tailbud Xenopus tropicalis cDNA clone TtBA070108 3',
mRNA sequence.
ACCESSION CR426638.1 GI:48920047
VERSION CR426638
KEYWORDS EST
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 652)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trops@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TtBA070108.g1kT7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7
Location/Qualifiers
1..652
/organism="Xenopus tropicalis"
/mol.type="mRNA"
/db.xref="taxon:8364"
/clone="TtBA070108"
/dev.stage="tailbud (stage 28-30)"
/lab.host="Escherichia coli DH10B."
/clone.lib="XGC-tailbud"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match      87.0%; Score 17.4; DB 7; Length 652;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAGAAACAG 19
|||||
Db 237 CAAGTAAATGAGAAACAG 255

RESULT 58
CN084914
LOCUS
DEFINITION EC2BBA24DC01.g1 Xenopus tropicalis xtbs plasmid library Xenopus
tropicalis cDNA clone xtbs24F02 5', mRNA sequence.
ACCESSION CN084914
VERSION CN084914.1 GI:45877610

```

```

EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 658)
Thuret,R., Fierro,A.C., Coen,L., Perron,M., Demeneix,B., Wegnez,M.,
Gaspay,G., Weissenbach,J., Wincker,P., Mazabraud,A. and Pollet,N.
Exploring the nervous system transcriptome in the model Xenopus
tropicalis using EST analysis
Unpublished (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Pollet N
Transgenese et Genetique des Amphibiens
CNRS UMR 8080
IBAIC bat 447, Universite Paris Sud, Orsay, F-91405, France
Tel: +33 169157272
Fax: +33 169156816
Email: Nicolas.Pollet@ibaic.u-psud.fr.
Location/Qualifiers
1..658
/organism="Xenopus tropicalis"
/mol.type="mRNA"
/strain="Ivory Coast"
/db.xref="taxon:8364"
/clone="xtbs24F02"
/tissue.type="pool of brains and spinal cords from
tadpoles at stages 51-52 and 61-62"
/dev.stage="stage 51-52 and 61-62"
/lab.host="E.coli DH10B"
/clone.lib="Xenopus tropicalis xtbs plasmid library"
/note="Vector: pCMVSPORT6 xtbs; Site_1: SfiI; Site_2:
SfiI; Xenopus tropicalis polyA+ RNA was obtained from
brain and spinal cord of tadpoles at stages 51-52 and
61-62. cDNAs were synthesized using the SMART system of
CLONTECH and directionally cloned into pCMVSPORT6 xtbs, a
modified version of pCMVSPORT6 allowing directional
cloning using asymmetric SfiI sites. For antisense RNA
synthesis, use T7 promoter and for sense RNA use SP6
promoter. Library constructed by Dr. L. Coen and Prof. B.
Demeneix (Museum National d'Histoire Naturelle and CNRS
UMR 5166, Paris, France)."

ORIGIN
Query Match      87.0%; Score 17.4; DB 7; Length 658;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAGAAACAG 19
|||||
Db 434 CAAGTAAATGAGAAACAG 416

RESULT 59
BJ076021
LOCUS
DEFINITION BJ076021 NIBB Mochii normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL104g09 3', mRNA sequence.
ACCESSION BJ076021
VERSION BJ076021.1 GI:17520937
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE
1 (bases 1 to 663)
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Tadaasu Shin-i
Center For Genetic Resource Information

```

National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshinigenes.nig.ac.jp  
The information of this clone is available through the following URL.  
<http://xenopus.nibb.ac.jp>.

FEATURES  
source  
Location/Qualifiers  
1..663  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL104g09"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"

ORIGIN  
Query Match 87.0%; Score 17.4; DB 3; Length 663;  
Best Local Similarity 94.7%; Pred. NO. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 197 CAAGTAAATGCAGAAACAG 215  
|||||

RESULT 60  
CX491676 669 bp mRNA linear EST 10-JAN-2005  
LOCUS JGI\_XZG38345.rev NIH\_XGC\_tropGas7 Xenopus tropicalis cDNA clone  
DEFINITION IMAGE:7553198 3', mRNA sequence.  
ACCESSION CX491676  
VERSION CX491676.1 GI:57403338  
KEYWORDS EST.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.  
REFERENCE 1 (bases 1 to 669)  
AUTHORS Richardson,P., Lucas S., Rokhsar,D., Dettter,J.C., Ng,D.C., Brokstein,P. and Lindquist,E.A.  
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project  
JOURNAL Unpublished (2004)  
COMMENT Other ESTs: JGI\_XZG38345.fwd  
Contact: Lindquist,E.A., Richardson,P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: [cdna@jgi-psf.org](mailto:cdna@jgi-psf.org)  
Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>  
cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley  
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LLNL: <http://image.llnl.gov>  
Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.  
Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.  
Plate: XZG 0397 row: b column: 12  
High quality sequence stop: 669.  
Location/Qualifiers  
1..669  
/organism="Xenopus tropicalis"

/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7553198"  
/tissue\_type="whole embryo"  
/dev\_stage="Gastrula (st. 10.5-12.5)"  
/lab\_host="E. coli XL1-Blue derivative, Stratagene Electropren-Blue"  
/clone\_lib="NIH\_XGC\_tropGas7"  
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Gastrula library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108  
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)"

ORIGIN  
Query Match 87.0%; Score 17.4; DB 8; Length 669;  
Best Local Similarity 94.7%; Pred. NO. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 232 CAAGTAAATGCAGAAACAG 250  
|||||

Search completed: January 28, 2006, 01:14:34  
Job time : 1628.9 secs

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 56.1798 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaacagg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/8B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCUS COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfileseq.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	17.4	87.0	160759	3	US-09-949-016-16514
2	16.8	84.0	2689	2	US-08-876-546A-15
3	16.8	84.0	2689	3	US-09-412-252-15
4	16.8	84.0	9633	9	5223423-1
5	16.8	84.0	9633	3	US-09-869-588-22
6	16.8	84.0	9726	3	US-09-869-588-21
7	16.8	84.0	24644	3	US-09-949-016-13113
8	16.4	82.0	601	3	US-09-949-016-10792
9	16.4	82.0	601	3	US-09-949-016-110815
10	16.4	82.0	601	3	US-09-949-016-110976
11	16.4	82.0	601	3	US-09-949-016-111137
12	16.4	82.0	601	3	US-09-949-016-111298
13	16.4	82.0	714	3	US-09-134-001C-2794
14	16.4	82.0	924	3	US-08-956-171E-745
15	16.4	82.0	924	3	US-08-781-986A-745
16	16.4	82.0	1035	2	US-07-941-523-22
17	16.4	82.0	1110	2	US-07-941-523-20
18	16.4	82.0	1196	3	US-09-489-847-15
19	16.4	82.0	1638	2	US-08-838-219B-8
20	16.4	82.0	1638	3	US-09-233-336A-8
21	16.4	82.0	1638	3	US-09-233-752A-8
22	16.4	82.0	1638	3	US-09-402-036-8
23	16.4	82.0	1638	3	US-09-904-226-8
24	16.4	82.0	21438	3	US-09-949-016-14198

C 25	16.4	82.0	28473	3	US-08-961-527-83	Sequence 83, Appl
C 26	16.4	82.0	89268	3	US-09-949-016-13025	Sequence 12025, A
C 27	16.4	82.0	90544	3	US-09-949-016-13302	Sequence 13302, A
C 28	16.4	82.0	171700	3	US-09-949-016-12276	Sequence 12276, A
C 29	16.4	82.0	171701	3	US-09-949-016-15835	Sequence 15835, A
C 30	16.4	82.0	174170	3	US-09-949-016-14810	Sequence 14810, A
C 31	16.4	82.0	174170	3	US-09-949-016-14811	Sequence 14811, A
C 32	16.4	82.0	174259	3	US-09-949-016-11968	Sequence 11968, A
C 33	16.4	82.0	174262	3	US-09-949-016-14259	Sequence 14259, A
C 34	16.4	82.0	174318	3	US-09-949-016-11880	Sequence 11880, A
C 35	16.4	82.0	174318	3	US-09-949-016-14812	Sequence 14812, A
C 36	16.4	82.0	174318	3	US-09-949-016-14813	Sequence 14813, A
C 37	16.4	82.0	390890	3	US-09-949-016-14720	Sequence 14720, A
C 38	16	80.0	601	3	US-09-949-016-68433	Sequence 68433, A
C 39	16	80.0	601	3	US-09-949-016-68434	Sequence 68434, A
C 40	16	80.0	601	3	US-09-949-016-68435	Sequence 68435, A
C 41	16	80.0	601	3	US-09-949-016-68436	Sequence 68436, A
C 42	16	80.0	95566	3	US-09-949-016-11877	Sequence 11877, A
C 43	16	80.0	112112	3	US-09-949-016-15639	Sequence 15639, A
C 44	16	80.0	130724	3	US-09-949-016-13753	Sequence 13753, A
C 45	15.8	79.0	210	3	US-08-847-065-15	Sequence 15, Appl
C 46	15.8	79.0	210	3	US-09-829-382-15	Sequence 15, Appl
C 47	15.8	79.0	510	3	US-09-583-110-2385	Sequence 2385, Ap
C 48	15.8	79.0	561	3	US-09-107-433-274	Sequence 274, App
C 49	15.8	79.0	601	3	US-09-949-016-66214	Sequence 66214, A
C 50	15.8	79.0	601	3	US-09-949-016-107420	Sequence 107420, A
C 51	15.8	79.0	601	3	US-09-949-016-107489	Sequence 107489, A
C 52	15.8	79.0	601	3	US-09-949-016-140767	Sequence 140767, A
C 53	15.8	79.0	601	3	US-09-949-016-140768	Sequence 140768, A
C 54	15.8	79.0	601	3	US-09-949-002-2533	Sequence 2533, Ap
C 55	15.8	79.0	601	3	US-09-949-002-2534	Sequence 2534, Ap
C 56	15.8	79.0	601	3	US-09-949-002-2535	Sequence 2535, Ap
C 57	15.8	79.0	601	3	US-09-949-002-4046	Sequence 4046, Ap
C 58	15.8	79.0	601	3	US-09-949-002-4047	Sequence 4047, Ap
C 59	15.8	79.0	601	3	US-09-949-002-4048	Sequence 4048, Ap
C 60	15.8	79.0	672	3	US-09-134-000C-2578	Sequence 2578, Ap

ALIGNMENTS

RESULT 1  
US-09-949-016-16514  
; Sequence 16514, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16514  
; LENGTH: 160759  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(160759)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16514

Query Match 87.0%; Score 17.4; DB 3; Length 160759;  
Best Local Similarity 94.7%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
Db 94050 CAAGTAAATTCAGAAACAG 94068

## RESULT 2

US-08-876-546A-15  
; Sequence 15, Application US/08876546A  
; Patent No. 5962665  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,546A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yaegeer  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER: 6127.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/938-3508  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic RNA (HIV-2)

US-08-876-546A-15  
Query Match 84.0%; Score 16.8; DB 2; Length 2689;  
Best Local Similarity 90.0%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158

RESULT 3  
US-09-412-252-15  
; Sequence 15, Application US/09412252  
; Patent No. 6232455  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING

QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158

## RESULT 3

US-09-412-252-15  
; Sequence 15, Application US/09412252  
; Patent No. 6232455  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING

; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/412,252  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,546  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yaegeer  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER: 6127.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/938-3508  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic RNA (HIV-2)

US-09-412-252-15  
Query Match 84.0%; Score 16.8; DB 3; Length 2689;  
Best Local Similarity 90.0%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158

RESULT 4  
5233423-1  
; Patent No. 5233423  
; APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAAL, FLOSSIE;  
; GALLO, ROBERT  
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT  
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/331,212  
; FILING DATE: 03-31-1989  
; SEQ ID NO:1:  
; LENGTH: 9633  
5233423-1

Query Match 84.0%; Score 16.8; DB 9; Length 9633;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 4209 CAAGTAAATGCAGAACTAGG 4228

## RESULT 5

US-09-869-588-22  
; Sequence 22, Application US/09869588



```
; Patent No. 6790657
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 59316
; CURRENT APPLICATION NUMBER: US/09/869,588
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; TYPE: DNA
; LENGTH: 9663
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PCM-ROD(SD36/EM)
; NAME/KEY: misc_feature
; LOCATION: (1)..(9663)
; OTHER INFORMATION: n represents a, c, g or t.
US-09-869-588-22

Query Match      84.0%; Score 16.8; DB 3; Length 9663;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
      |||||
Db      4818 CAAGTAAATGCAGAACTAGG 4837

RESULT 6
US-09-869-588-21
; Sequence 21, Application US/09869588
; Patent No. 6790657
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 59316
; CURRENT APPLICATION NUMBER: US/09/869,588
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 9726
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PROD(SD36/EM)
; NAME/KEY: misc_feature
; LOCATION: (1)..(9726)
; OTHER INFORMATION: n represents a, c, t, or g.
US-09-869-588-21

Query Match      84.0%; Score 16.8; DB 3; Length 9726;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
      |||||
Db      4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 7
US-09-949-016-13113/c
; Sequence 13113, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13113
; LENGTH: 246444
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)....(246444)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13113

Query Match      84.0%; Score 16.8; DB 3; Length 246444;
Best Local Similarity 90.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
      |||||
Db      36561 CAAGTAAAGGCAGAAATAGG 36542

RESULT 8
US-09-949-016-107692
; Sequence 107692, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107692
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)....(601)
; OTHER INFORMATION: n represents a, c, t, or g.
US-09-949-016-107692

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  AAGTAAATGCAGAAACAG 19
      |||||
Db      396 AAGAAATGCAGAAACAG 413

RESULT 9
US-09-949-016-110815/c
; Sequence 110815, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110815
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-110815

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 10
US-09-949-016-110976/c
; Sequence 110976, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110976
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-110976

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 11
US-09-949-016-111137/c
; Sequence 111137, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111137
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-111137

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 12
US-09-949-016-111298/c
; Sequence 111298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111298
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-111298

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 13
US-09-134-001C-2794/c
; Sequence 2794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
```

;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 2794  
;; LENGTH: 714  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2794

Query Match 82.0%; Score 16.4; DB 3; Length 714;  
Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACA 18  
Db 137 CAAGTAAATGCAGAAACA 120  
|||||

RESULT 14  
US-08-956-171E-745  
; Sequence 745, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hyman  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 745:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 924 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 745:  
US-08-956-171E-745

Query Match 82.0%; Score 16.4; DB 3; Length 924;  
Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACA 18  
Db 137 CAAGTAAATGCAGAAACA 120  
|||||

RESULT 15  
US-08-781-986A-745  
; Sequence 745, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 745:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 924 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-745

Query Match 82.0%; Score 16.4; DB 3; Length 924;  
Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACA 18  
Db 273 CAAGTAAATGCAGAAACA 290  
|||||

RESULT 16  
US-07-941-523-22  
; Sequence 22, Application US/07941523  
; Patent No. 5571718  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John J  
; APPLICANT: Barbour, Alan G  
; TITLE OF INVENTION: Cloning and Expression of Borrelia  
; TITLE OF INVENTION: Lipoproteins  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 01730  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,523
; FILING DATE: 19920908
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: BNL90-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1035 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-941-523-22

Query Match      82.0%; Score 16.4; DB 2; Length 1035;
Best Local Similarity 94.4%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
Db      246 AGTAAAGCAGAAACAGG 263

RESULT 17
US-07-941-523-40
; Sequence 20, Application US/07941523
; Patent No. 5571718
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J
; APPLICANT: Barbour, Alan G
; TITLE OF INVENTION: Cloning and Expression of Borrelia
; TITLE OF INVENTION: Lipoproteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 01730
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,523
; FILING DATE: 19920908
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: BNL90-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-941-523-20
```

```
Query Match      82.0%; Score 16.4; DB 2; Length 1110;
Best Local Similarity 94.4%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
Db      321 AGTAAAGCAGAAACAGG 338

RESULT 18
US-09-489-847-15/c
; Sequence 15, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-489-847-15
```

```
Query Match      82.0%; Score 16.4; DB 3; Length 1196;
Best Local Similarity 94.4%; Pred. No. 2.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACA 18
Db      311 CAAATAAATGCAGAAACA 294

RESULT 19
US-08-838-219B-8
; Sequence 8, Application US/08838219B
; Patent No. 5877012
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: A No. 5877012el Class of Proteins for the
; TITLE OF INVENTION: Control of Plant Pests
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/838,219B  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA  
encoding VIP3A(a) receptor"  
US-08-838-219B-8

Query Match 82.0%; Score 16.4; DB 2; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 334 AAGTAGTGCAGAAACAG 351

RESULT 20  
US-09-233-336A-8  
Sequence 8, Application US/09233336A  
Patent No. 6107279  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6107279el Class of Proteins for the  
CONTROL OF PLANT PESTS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne

STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233,336A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA  
encoding VIP3A(a) receptor"  
US-09-233-336A-8

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 334 AAGTAGTGCAGAAACAG 351

RESULT 21  
US-09-233-752A-8  
Sequence 8, Application US/09233752A  
Patent No. 6137033  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the  
CONTROL OF PLANT PESTS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233,752A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA encoding VIP3A(a) receptor"  
US-09-233-752A-8

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
Db 334 AAGTAGATGCAGAAACAG 351

RESULT 22  
US-09-402-036-8  
Sequence 8, Application US/09402036  
Patent No. 6291156  
GENERAL INFORMATION:  
APPLICANT: Estruch, Juan J.  
APPLICANT: Yu, Cao-Guo  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Koziel, Michael  
APPLICANT: Nye, Gordon  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284C  
CURRENT APPLICATION NUMBER: US/09/402,036  
CURRENT FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: PCT/EP98/01952  
PRIOR FILING DATE: 1998-04-02  
PRIOR APPLICATION NUMBER: 08/838,219  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,263  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,265  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/463,483  
PRIOR FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/314,594  
PRIOR FILING DATE: 1994-09-09  
PRIOR APPLICATION NUMBER: 08/218,018  
PRIOR FILING DATE: 1994-03-24  
PRIOR APPLICATION NUMBER: 08/037,057  
PRIOR FILING DATE: 1993-03-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1638  
TYPE: DNA  
ORGANISM: Agrotis ipsilon  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)..(1189)  
OTHER INFORMATION: Translation of cDNA encoding VIP3A(a) receptor  
US-09-402-036-8

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
Db 334 AAGTAGATGCAGAAACAG 351

RESULT 23  
US-09-904-226-8  
Sequence 8, Application US/09904226  
Patent No. 6429360  
GENERAL INFORMATION:  
APPLICANT: Estruch, Juan J.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Koziel, Michael  
APPLICANT: Nye, Gordon  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284D  
CURRENT APPLICATION NUMBER: US/09/904,226  
CURRENT FILING DATE: 2001-07-12  
PRIOR APPLICATION NUMBER: PCT/EP98/01952  
PRIOR FILING DATE: 1998-04-02  
PRIOR APPLICATION NUMBER: 08/838,219  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,263  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,265  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/463,483  
PRIOR FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/314,594  
PRIOR FILING DATE: 1994-09-09  
PRIOR APPLICATION NUMBER: 08/218,018  
PRIOR FILING DATE: 1994-03-24  
PRIOR APPLICATION NUMBER: 08/037,057  
PRIOR FILING DATE: 1993-03-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1638  
TYPE: DNA

```
; ORGANISM: Agrotis ipsilon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1189)
; OTHER INFORMATION: Translation of cDNA encoding VIP3A(a) receptor
; OTHER INFORMATION: from Black cutworm
US-09-904-226-8

Query Match      82.0%; Score 16.4; DB 3; Length 1638;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      334 AAGTAGATGCAGAAACAG 351

RESULT 24
US-09-949-016-14198/c
; Sequence 14198, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14198
; LENGTH: 21438
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14198

Query Match      82.0%; Score 16.4; DB 3; Length 21438;
Best Local Similarity 94.4%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      14869 AAGTAAATGAAGAAACAG 14852

RESULT 25
US-08-961-527-83/c
; Sequence 83, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
```

```
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-83

Query Match      82.0%; Score 16.4; DB 3; Length 28473;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      16919 AAGTAAATGCAGTAAACAG 16902

RESULT 26
US-09-949-016-12025/c
; Sequence 12025, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12025
; LENGTH: 89268
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12025

Query Match      82.0%; Score 16.4; DB 3; Length 89268;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      33178 AAGTAAAGGCAGAAACAG 33161

RESULT 27
US-09-949-016-13302/c
; Sequence 13302, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13302
; LENGTH: 90544
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13302

Query Match      82.0%; Score 16.4; DB 3; Length 90544;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 33363 AAGTAAAGCGAGAAACAG 33346

RESULT 28
US-09-949-016-12276/c
; Sequence 12276, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12276
; LENGTH: 171700
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12276

Query Match      82.0%; Score 16.4; DB 3; Length 171700;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 18
    ||||| ||||| ||||| |||||
Db 45460 CAAGTAAATGCATAAACA 45443

RESULT 29
US-09-949-016-15835/c
; Sequence 15835, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15835
; LENGTH: 171700
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15835

Query Match      82.0%; Score 16.4; DB 3; Length 171700;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 94124 AAGTAAATGCAGAAACAG 94107

RESULT 30
US-09-949-016-14810/c
; Sequence 14810, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14810
; LENGTH: 174170
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174170)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14810

Query Match      82.0%; Score 16.4; DB 3; Length 174170;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 94124 AAGTAAATGCAGAAACAG 94107

RESULT 31
US-09-949-016-14811/c
; Sequence 14811, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14811
; LENGTH: 174170
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174170)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14811

Query Match      82.0%; Score 16.4; DB 3; Length 174170;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 94124 AAGTAAATGCAGAAACAG 94107
```



; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14811  
; LENGTH: 174170  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)-(174170)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14811

Query Match 82.0%; Score 16.4; DB 3; Length 174170;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 94124 AAGTAAATGCAGAAACAG 94107

## RESULT 32

US-09-949-016-11968  
; Sequence 11968, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11968  
; LENGTH: 174259  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11968

Query Match 82.0%; Score 16.4; DB 3; Length 174259;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 105890 AAGTAAATGCAGAAACAG 105907

## RESULT 33

US-09-949-016-14259  
; Sequence 14259, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14259  
; LENGTH: 174262  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14259

Query Match 82.0%; Score 16.4; DB 3; Length 174262;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 105890 AAGTAAATGCAGAAACAG 105907

## RESULT 34

US-09-949-016-11880/c  
; Sequence 11880, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11880  
; LENGTH: 174318  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(174318)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-11880

Query Match 82.0%; Score 16.4; DB 3; Length 174318;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 94272 AAGTAAATGCAGAAACAG 94255

## RESULT 35

US-09-949-016-14812/c  
; Sequence 14812, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012



Mon Jan 30 07:42:54 2006

; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68434  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68434

Query Match 80.0%; Score 16; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||  
Db 239 TAAATGCAGAAACAGG 224

RESULT 40  
US-09-949-016-68435/c  
; Sequence 68435, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68435  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68435

Query Match 80.0%; Score 16; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||  
Db 465 TAAATGCAGAAACAGG 450

RESULT 41  
US-09-949-016-68436/c  
; Sequence 68436, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68436  
; LENGTH: 601

; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68436

Query Match 80.0%; Score 16; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||  
Db 486 TAAATGCAGAAACAGG 471

RESULT 42  
US-09-949-016-11877  
; Sequence 11877, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11877  
; LENGTH: 95566  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11877

Query Match 80.0%; Score 16; DB 3; Length 95566;  
Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GTAAATGCAGAAACAG 19  
|||||  
Db 14408 GTAAATGCAGAAACAG 14423

RESULT 43  
US-09-949-016-15639/c  
; Sequence 15639, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15639  
; LENGTH: 112112  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15639

Query Match 80.0%; Score 16; DB 3; Length 112112;

```
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACA 18
Db 5625 AGTAAATGCAGAAACA 5610

RESULT 44
US-09-949-016-13753
; Sequence 13753, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13753
; LENGTH: 130724
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(130724)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13753

Query Match 80.0%; Score 16; DB 3; Length 130724;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20
Db 8750 TAAATGCAGAAACAGG 8765

RESULT 45
US-08-847-065-15
; Sequence 15, Application US/08847065
; Patent No. 6245335
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosenow, Carsten I.
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Wizemann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,065
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
```

```
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: B. subtilis
US-08-847-065-15

Query Match 79.0%; Score 15.8; DB 3; Length 210;
Best Local Similarity 89.5%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20
Db 29 AAGTATATACAGAAACAGG 47

RESULT 46
US-09-829-382-15
; Sequence 15, Application US/09829382
; Patent No. 6784164
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosenow, Carsten I.
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Wizemann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/829,382
; FILING DATE: 09-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/847,065
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
```

```

> OPERATING SYSTEM: <Unknown>
> SOFTWARE: <Unknown>
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/09/107,433
> FILING DATE: 30-Jun-1998
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 60/ 085131
> FILING DATE: May 12, 1998
> APPLICATION NUMBER: 60/051553
> FILING DATE: July 2, 1997
> ATTORNEY/AGENT INFORMATION:
> NAME: Ariniello, Pamela Deneke
> REGISTRATION NUMBER: 40,489
> REFERENCE/DOCKET NUMBER: GTC-011
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (781)893-5007
> TELEFAX: (781)893-8277
> INFORMATION FOR SEQ ID NO: 274:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 561 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: circular
> MOLECULE TYPE: DNA (genomic)
> HYPOTHETICAL: NO
> ANTI-SENSE: NO
> ORIGINAL SOURCE:
> ORGANISM: Streptococcus pneumoniae
> FEATURE:
> NAME/KEY: misc feature
> LOCATION: (B) LOCATION 1...561
> SEQUENCE DESCRIPTION: SEQ ID NO: 274:
>
> US-09-107-433-274
>
> Query Match 79.0%; Score 15.8; DB 3; Length 561;
> Best Local Similarity 89.5%; Pred. No. 4.8e+02;
> Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
>
> QY 1 CAAGTAAATGCAGAAACAG 19
> ||| ||||| ||||| |||||
> Db 553 CAATTAAATGCAGTAACAG 535
>
> RESULT 49
> US-09-949-016-66214
> Sequence 66214, Application US/09949016
> Patent No. 6812339
> GENERAL INFORMATION:
> APPLICANT: VENTER, J. Craig et al.
> TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
> WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
> FILE REFERENCE: CL001307
> CURRENT APPLICATION NUMBER: US/09/949,016
> CURRENT FILING DATE: 2000-04-14
> PRIOR APPLICATION NUMBER: 60/241,755
> PRIOR FILING DATE: 2000-10-20
> PRIOR APPLICATION NUMBER: 60/237,768
> PRIOR FILING DATE: 2000-10-03
> PRIOR APPLICATION NUMBER: 60/231,498
> PRIOR FILING DATE: 2000-09-08
> NUMBER OF SEQ ID NOS: 207012
> SOFTWARE: FastSeq for Windows Version 4.0
> SEQ ID NO 66214
> LENGTH: 601
> TYPE: DNA
> ORGANISM: Human
>
> US-09-949-016-66214
>
> Query Match 79.0%; Score 15.8; DB 3; Length 601;
> Best Local Similarity 89.5%; Pred. No. 4.8e+02;
> Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
>
> QY 1 CAAGTAAATGCAGAAACAG 19

```

RESULT 52

RESULT 54  
US-09-949-002-2533/c  
; Sequence 2533, Application US/09949002  
; Patent No. 6900016  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2533
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2533

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| |||||||||||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 55
US-09-949-002-2534/c
; Sequence 2534, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2534
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2534

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| |||||||||||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 56
US-09-949-002-2535/c
; Sequence 2535, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2535
```

```
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2535

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| |||||||||||||
Db 82 CAAATTAATGCAGAAACAG 64

RESULT 57
US-09-949-002-4046/c
; Sequence 4046, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4046
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-4046

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| |||||||||||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 58
US-09-949-002-4047/c
; Sequence 4047, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4047
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-4047

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| |||||||||||||
```

Db 143 CAAATTATGCAGAAACAG 125

## RESULT 59

US-09-949-002-4048/c  
; Sequence 4048, Application US/09949002  
; Patent No. 6900016

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

1. TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION

1. TITLE OF INVENTION: WITH INTRODUCTION

; TITLE OF INVENTION: AND  
: FILE REFERENCE: CL000790FILE REFERENCE: CF000750  
CURRENT APPLICATION NUMBER: US/09/949.002

: CURRENT FILING DATE: 2000-01-28  
 : CURRENT AFFILIATION NUMBER: 03/0

; CURRENT FILING DATE: 2000-01-28  
 : PRIOR APPLICATION NUMBER: 60/231,407

; PRIOR AFFILIATION NUMBER: 6007  
; PRIOR FILING DATE: 2000-09-08

; PRIOR FILING DATE: 2000-09-  
 : NUMBER OF SEQ ID NOS: 10823

```

; NUMBER OF SEQ ID NOS: 10823
: SOFTWARE: FastSeq for Windows Version 4.0

```

; SOFTWARE: FABO  
; CEO ID NO 1048

```

; SEQ ID NO 404
: I ENCTH: 601

```

```

; LENGTH: 60
TYPE: DNA

```

TYPE: DNA

; ORGANISM: Human  
US-09-949-002-4048

```
Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

1 CAAGTAAATGCAGAAACAG 19

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

## RESULT 60

US-09-134-000C-2578  
; Sequence 2578, Application US/09134000C  
; Patent No. 6617156

: GENERAL INFORMATION:

: **APPLICANT:** Lynn Doucette-Stamm et al

RELEVANT: *LYNN DOUGLASS-SEARNS* ET AL  
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

1. TITLE OF INVENTION: NOCETEIC ACID AND AMINO ACID SEQUENCES RETAINING 10  
2. TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE OF INVENTION: ENYER;  
FILE REFERENCE: 032796-032

FILE REFERENCE: 052756-032  
: CURRENT APPLICATION NUMBER: IIS/09/134.000C

CURRENT AFFILIATION NUMBER: 09/0  
: CURRENT FILING DATE: 1998-08-13

; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/055,778

PRIOF AFFILIATION NUMBER: US  
: PRIOF FILING DATE: 1997-08-15

: PRIOR FILING DATE: 1997-08  
 : NUMBER OF SEQ ID NOS: 6812

```

; NUMBER OF SEQ ID NOS: 6812
: SOFTWARE: Patent version 3.1

```

; SOFTWARE: PATCO  
; CEO ID NO 3578

; SEQ ID NO 257

LENGTH: 67

TYPE: DNA

; ORGANISM: Enterobacteriaceae;

Query Match 79.0%; Score 15.8; DB 3; Length 672;  
Best Local Similarity 89.5%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 AAGTAAATGCAGAAACAGG 20

[illegible]

Search completed: January 28, 2006, 01:18:50  
Job time : 62.1798 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 346.966 Seconds  
(without alignments) 476.668 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaatgcagaaacagg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum DB seq length:	0
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

FOBC-processing: Minimum Match 0%  
Maximum Match 100%

Maximum value 200  
Listing first 60 summaries

```

Database : Published Applications NA Main:
1: /cgn2_6/prodata/1/pubnpa/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubnpa/US08_PUBCOMB.seq.*
3: /cgn2_6/prodata/1/pubnpa/US09A_PUBCOMB.seq.*
4: /cgn2_6/prodata/1/pubnpa/US09B_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubnpa/US10A_PUBCOMB.seq.*
6: /cgn2_6/prodata/1/pubnpa/US10B_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubnpa/US10C_PUBCOMB.seq.*
8: /cgn2_6/prodata/1/pubnpa/US10D_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubnpa/US10E_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubnpa/US11_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SIMMARTES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	20	100.0	20	9	US-10-716-005-4	Sequence 4, Appli
C 2	17.4	87.0	272	3	US-09-815-242-144	Sequence 144, App
C 3	17.4	87.0	272	7	US-10-282-130	Sequence 130, App
4	17.4	87.0	442	5	US-10-027-632-193671	Sequence 193671,
5	17.4	87.0	442	6	US-10-027-632-193671	Sequence 193671,
C 6	17.4	87.0	543	4	US-09-925-065A-767188	Sequence 767188,
7	17.4	87.0	732	7	US-10-282-132A-6267	Sequence 6267, Ap
C 8	17.4	87.0	4911	10	US-11-097-143-13502	Sequence 13502, A
C 9	17.4	87.0	7886	10	US-11-097-143-13501	Sequence 13501, A
10	17.4	87.0	95090	8	US-10-656-394A-13	Sequence 13, Appl
11	17.4	87.0	177587	5	US-10-087-192-1438	Sequence 1438, Ap
12	17	85.0	603	4	US-09-925-065A-332186	Sequence 332186,
13	17	85.0	603	4	US-09-925-065A-332187	Sequence 332187,
C 14	17	85.0	703	5	US-10-027-632-127740	Sequence 127740,
C 15	17	85.0	703	6	US-10-027-632-127740	Sequence 127740,
C 16	17	85.0	1081	3	US-09-070-927A-773	Sequence 773, App
17	16.8	84.0	426	4	US-09-925-065A-975556	Sequence 975556, A
18	16.8	84.0	426	4	US-09-925-065A-136681	Sequence 136681,
19	16.8	84.0	426	4	US-09-925-065A-136682	Sequence 136682,
20	16.8	84.0	445	4	US-09-925-065A-364908	Sequence 364908,
21	16.8	84.0	565	4	US-09-925-065A-614612	Sequence 614612,
22	16.8	84.0	583	4	US-09-925-065A-429130	Sequence 429130,
23	16.8	84.0	583	4	US-09-925-065A-429131	Sequence 429131,

## ALIGNMENTS

## RESULT 1

US-10-716-005-4

US-10-718-003-1  
: Sequence 4. Application US/10716005

; Publication No. US20050106578A1

GENERAL INFORMATION:

APPLICANT: Uhl, James R.

APPLICANT: Cockerill III, Frank

APPLICANT: Aichinger, Christian

; APPLICANT: Reiser, Astrid

; TITLE OF INVENTION: Detection of Group B

FILE REFERENCE: 07039/460001

; CURRENT APPLICATION NUMBER: US/1

; CURRENT FILING DATE: 2010-01-26

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: F

; SEQ ID NO 4

; LENGTH: 20  
TYPE: DIA

TYPE: DNA  
ORGANISM: Artificial Sequence

ORGANISM: Artificial Sequence

```
; ; FEATURE: oligonucleotide
```

; OTHER INFO  
IIS-10-716-005-4

Query	Match	100 0%:	Score 20:	DB 9:	Length 20:
-------	-------	---------	-----------	-------	------------

Query Match	100.0%;	score 20;	best
Best Local Similarity	100.0%;	Pred No. 22;	

BEST LOCAL SIMILARITY 100.0%; FREQ: NO: 22,  
Matches 20: Conservative 0: Mismatches 0: Indels 0:

1 CAAGTAAATGCAGAAACAGG 20

QY I CAAGTAAATGCAGAAACAGG 20  
|||||

```
RESULT 2
US-09-815-242-144/c
; Sequence 144, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-144

Query Match      87.0%; Score 17.4; DB 3; Length 272;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAATGCGAAGAACAG 19
      |||||
Db      166 CAAGTAATGAGAGAACAG 148

RESULT 3
US-10-282-122A-130/c
; Sequence 130, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

Query Match      87.0%; Score 17.4; DB 5; Length 442;
Best Local Similarity 94.7%; Pred. No. 5.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAATGCGAAGAACAG 19
      |||||
Db      166 CAAGTAATGAGAGAACAG 148

RESULT 4
US-10-027-632-193671
; Sequence 193671, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193671
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-193671

Query Match      87.0%; Score 17.4; DB 5; Length 442;
Best Local Similarity 94.7%; Pred. No. 5.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAATGCGAAGAACAG 19
      |||||
Db      166 CAAGTAATGAGAGAACAG 148

US-10-282-122A-130
Query Match      87.0%; Score 17.4; DB 7; Length 272;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAATGCGAAGAACAG 19
      |||||
Db      166 CAAGTAATGAGAGAACAG 148
```

QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
Db 247 AAGTAAATACAGAAACAGG 265

## RESULT 5

US-10-027-632-193671  
; Sequence 193671, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193671  
; LENGTH: 442  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-193671

Query Match 87.0%; Score 17.4; DB 6; Length 442;  
Best Local Similarity 94.7%; Pred. No. 5.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
Db 247 AAGTAAATACAGAAACAGG 265

## RESULT 6

US-09-925-065A-767188/c  
; Sequence 767188, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 767188  
; LENGTH: 543  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-925-065A-767188

Query Match 87.0%; Score 17.4; DB 4; Length 543;  
Best Local Similarity 94.7%; Pred. No. 5.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
Db 530 AAGAAATGCAGAAACAGG 512

## RESULT 7

US-10-282-122A-6267  
; Sequence 6267, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6267  
; LENGTH: 732  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis

US-10-282-122A-6267

Query Match 87.0%; Score 17.4; DB 7; Length 732;  
Best Local Similarity 94.7%; Pred. No. 6.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
||||||| |||||||  
Db 118 CAAGTAAATGAAGAAACAG 136

## RESULT 8

US-11-097-143-13502/c

```
; Sequence 13502, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13502
; LENGTH: 4911
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-13502

Query Match      87.0%; Score 17.4; DB 10; Length 4911;
Best Local Similarity 94.7%; Pred. No. 8.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      1405 AGGTAATGCAGAAACAGG 1387

RESULT 9
US-11-097-143-13501/c
; Sequence 13501, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23

; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13501
; LENGTH: 7886
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-13501

Query Match      87.0%; Score 17.4; DB 10; Length 7886;
Best Local Similarity 94.7%; Pred. No. 9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      3324 AGGTAATGCAGAAACAGG 3306

RESULT 10
US-10-656-394A-13
; Sequence 13, Application US/10656394A
; Publication No. US20040210957A1
; GENERAL INFORMATION:
; APPLICANT: Wang et al.
; TITLE OF INVENTION: Cloning and Characterization of the
; TITLE OF INVENTION: broad-spectrum resistance gene Pi2
; FILE REFERENCE: 035718/252062
; CURRENT APPLICATION NUMBER: US/10/656,394A
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 99090
; TYPE: DNA
; ORGANISM: Oryza minuta
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 23216
; OTHER INFORMATION: n = A,T,C or G
US-10-656-394A-13

Query Match      87.0%; Score 17.4; DB 8; Length 99090;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      32637 AAGTAAATGCAGAAACAGG 32655

RESULT 11
US-10-087-192-1438
; Sequence 1438, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1438
; LENGTH: 177587
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(177587)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1438

Query Match      87.0%; Score 17.4; DB 5; Length 177587;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   |||||
Db 102050 CAAGTAAATGCAGAAACAG 102068

RESULT 12
US-09-925-065A-332186
; Sequence 332186, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332186
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-332186

Query Match      85.0%; Score 17; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   |||||
Db 172 AGTAAATGCAGAAACAG 188

RESULT 13
US-09-925-065A-332187
; Sequence 332187, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332187
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-332187

Query Match      85.0%; Score 17; DB 5; Length 703;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACA 18
   |||||
Db 196 AAGTAAATGCAGAAACA 180

RESULT 15
US-10-027-632-127740/c
; Sequence 127740, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127740
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-127740

Query Match      85.0%; Score 17; DB 6; Length 703;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2  AAGTAAATGCAGAAACA 18
        |||||
Db      196 AAGTAAATGCAGAAACA 180

RESULT 16
US-09-070-927A-773/c
; Sequence 773, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;              Patrick J. Dillon
;              Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 773:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1081 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 773:
US-09-070-927A-773

Query Match      85.0%; Score 17; DB 3; Length 1081;
Best Local Similarity 89.5%; Pred. No. 1e+03;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1  CAAGTAAATGCAGAAACAG 19
        |||||
Db      608 CAAGTAAATGARGAAACAG 590

RESULT 17
US-09-925-065A-97556
; Sequence 97556, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97556
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-97556

Query Match      84.0%; Score 16.8; DB 4; Length 426;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  CAAGTAAATGCAGAAACAG 20
        |||||
Db      246 CAAGTAAATGCAGAAACAG 265

RESULT 18
US-09-925-065A-136681
; Sequence 136681, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 136681
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-136681

Query Match      84.0%; Score 16.8; DB 4; Length 426;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 246 CAAGTAAATGCAGAAACAGG 265

RESULT 19
US-09-925-065A-136682
; Sequence 136682, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136682
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-136682

Query Match      84.0%; Score 16.8; DB 4; Length 426;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 246 CAAGTAAATGCAGAAACAGG 265

RESULT 20
US-09-925-065A-364908
; Sequence 364908, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

---

```
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 364908
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-364908

Query Match      84.0%; Score 16.8; DB 4; Length 445;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 256 CAAGTAAATGCAGAAACAGG 275

RESULT 21
US-09-925-065A-614612
; Sequence 614612, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 614612
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-614612

Query Match      84.0%; Score 16.8; DB 4; Length 565;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 480 CAAGTAAAGGCAGAAATAGG 499

RESULT 22
US-09-925-065A-429130
; Sequence 429130, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

us-10-716-005-4.rnpbm

Mon Jan 30 07:42:55 2006

```

; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429130
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-429130

Query Match      84.0%; Score 16.8; DB 4; Length 583;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 398 CAAGAAAATGCAGAAAACAGG 417

RESULT 23
US-09-925-065A-429131
; Sequence 429131, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429131
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-429131

Query Match      84.0%; Score 16.8; DB 4; Length 583;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 398 CAAGAAAATGCAGAAAACAGG 417

RESULT 24
US-09-925-065A-259716
; Sequence 259716, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429131
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-429131

Query Match      84.0%; Score 16.8; DB 4; Length 583;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 398 CAAGAAAATGCAGAAAACAGG 417

RESULT 25
US-09-925-065A-45434
; Sequence 45434, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45434
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-45434

Query Match      84.0%; Score 16.8; DB 4; Length 614;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 327 CAAGTAACTGCAGACACAGG 346

RESULT 26
US-10-027-632-289556
; Sequence 289556, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
```



**Qy**            1 CAAAGTAAATGCAGAACAGG 20  
               ||| ||||| ||||| ||||| ||  
**Dβ**            442 CATGTAAATGCAGAACATG 461

Query Match 84.0%; Score 16.8; DB 5; Length 692;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;

```
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 248 CAAGAAATGCAGAAATCAGG 267

RESULT 30
US-10-027-632-101447
; Sequence 101447, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101447
; LENGTH: 692
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101447

Query Match 84.0%; Score 16.8; DB 6; Length 692;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 248 CAAGAAATGCAGAAATCAGG 267

RESULT 31
US-10-425-115-161074/c
; Sequence 161074, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 161074
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(821)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: MRT4577_78481C.1
US-10-425-115-161074

Query Match 84.0%; Score 16.8; DB 8; Length 821;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 627 CAAGAAATGCAGAAACAGG 608

RESULT 32
US-09-925-065A-550161
; Sequence 550161, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 550161
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-550161

Query Match 84.0%; Score 16.8; DB 4; Length 843;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 721 CAAGTAAATGCAGAAACAGG 740

RESULT 33
US-09-925-065A-550162
; Sequence 550162, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 550162
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-550162

Query Match      84.0%; Score 16.8; DB 4; Length 843;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 721 CAAGTGAATGCAAAAACAGG 740

RESULT 34
US-10-437-963-26351
; Sequence 26351, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 26351
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3114C.1
US-10-437-963-26351

Query Match      84.0%; Score 16.8; DB 7; Length 1812;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 451 CAAGCAATGCAGCAACAGG 470

RESULT 35
US-10-731-988-22
; Sequence 22, Application US/10731988
; Publication No. US20040147026A1
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 67517
; CURRENT APPLICATION NUMBER: US/10/731,988
; CURRENT FILING DATE: 2003-12-09
; PRIOR APPLICATION NUMBER: 09/869,588
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9663
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PCM-ROD(SD36/EM)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(9663)
; OTHER INFORMATION: n represents a, c, g or t.
US-10-731-988-22

Query Match      84.0%; Score 16.8; DB 7; Length 9663;
Best Local Similarity 90.0%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 4818 CAAGTAAATGCAGAACTAGG 4837

RESULT 36
US-10-731-988-21
; Sequence 21, Application US/10731988
; Publication No. US20040147026A1
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 67517
; CURRENT APPLICATION NUMBER: US/10/731,988
; CURRENT FILING DATE: 2003-12-09
; PRIOR APPLICATION NUMBER: 09/869,588
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 9726
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PROD(SD36/EM)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(9726)
; OTHER INFORMATION: n represents a, c, t, or g.
US-10-731-988-21

Query Match      84.0%; Score 16.8; DB 7; Length 9726;
Best Local Similarity 90.0%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 37
US-10-017-161-1989/c
; Sequence 1989, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABEURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1989
; LENGTH: 14241
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(14241)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(267)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4723)..(4988)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5026)..(5670)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13871)..(14041)
; LOCATION: (13871)..(14041)
US-10-017-161-1989

Query Match      84.0%; Score 16.8; DB 6; Length 14241;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 4047 CAAGTAAATGCAGAAACAGG 4028

RESULT 38
US-10-292-798-1637/c
; Sequence 1637, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABIURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1637
; LENGTH: 14425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; LOCATION: (1)..(14425)
; NAME/KEY: CDS
; LOCATION: (201)..(235)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (306)..(436)
; NAME/KEY: CDS
; LOCATION: (4906)..(5171)
; NAME/KEY: CDS
; LOCATION: (5209)..(5853)
; NAME/KEY: CDS
; LOCATION: (14055)..(14225)

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1989
; LENGTH: 14241
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(14241)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(267)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4723)..(4988)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5026)..(5670)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13871)..(14041)
; LOCATION: (13871)..(14041)
US-10-017-161-1989

Query Match      84.0%; Score 16.8; DB 6; Length 14241;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 4047 CAAGTAAATGCAGAAACAGG 4028

RESULT 39
US-10-017-161-2251/c
; Sequence 2251, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABIURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2251
; LENGTH: 16270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(16270)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(286)
; NAME/KEY: CDS
; LOCATION: (2151)..(2281)
; NAME/KEY: CDS
; LOCATION: (6751)..(7016)
; NAME/KEY: CDS
; LOCATION: (7054)..(7698)
; NAME/KEY: CDS
; LOCATION: (15900)..(16070)
US-10-017-161-2251

Query Match      84.0%; Score 16.8; DB 6; Length 16270;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 6075 CAAGTAAATGCAGAAACAGG 6056

RESULT 40
US-10-292-798-1897/c
; Sequence 1897, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABIURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
```

; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1897  
; LENGTH: 16270  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; LOCATION: source  
; FEATURE:  
; LOCATION: (1)..(16270)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (201)..(286)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2151)..(2281)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (6751)..(7016)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7054)..(7698)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (15900)..(16070)  
; US-10-292-798-1897

Query Match 84.0%; Score 16.8; DB 6; Length 16270;  
Best Local Similarity 90.0%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
Db 6075 CAAGTAAATGCAGAAACAGG 6056

RESULT 41  
US-10-322-281-516  
; Sequence 516, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 516  
; LENGTH: 41079  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(41079)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-322-281-516

Query Match 84.0%; Score 16.8; DB 7; Length 41079;  
Best Local Similarity 90.0%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
Db 2956 CATGTAATGCAGAAACATG 2975

RESULT 42  
US-10-719-993-6799/c  
; Sequence 6799, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6799  
; LENGTH: 255439  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-719-993-6799

Query Match 84.0%; Score 16.8; DB 8; Length 255439;  
Best Local Similarity 90.0%; Pred. No. 2.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
Db 40511 CAAGTAAAGGCAGAAATAGG 40492

RESULT 43  
US-10-424-599-27588/c  
; Sequence 27588, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovacic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 27588  
; LENGTH: 275  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(275)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_124913C.1  
US-10-424-599-27588

Query Match 82.0%; Score 16.4; DB 7; Length 275;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
Db 29 AATTAATGCAGAAACAG 12

RESULT 44  
US-09-925-065A-168362  
; Sequence 168362, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135

A1

```
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 653327
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-653327

Query Match      82.0%; Score 16.4; DB 4; Length 458;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
      ||||| ||||| ||||| |||||
Db      439 AGTAAATGCAGAAACAGG 456

RESULT 49
US-09-925-065A-222603/c
; Sequence 222603, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 222603
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-222603

Query Match      82.0%; Score 16.4; DB 4; Length 464;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
      ||||| ||||| ||||| |||||
Db      20 AGTAAATGCAGAAACAGG 3

RESULT 50
US-09-925-065A-150439/c
; Sequence 150439, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150439
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-150439

Query Match      82.0%; Score 16.4; DB 4; Length 503;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      42 AAGTAAATGCAGAAACAG 25

RESULT 51
US-09-925-065A-150440/c
; Sequence 150440, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150440
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-150440

Query Match      82.0%; Score 16.4; DB 4; Length 503;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      42 AAGTAAATGCAGAAACAG 25

RESULT 52
US-10-027-632-6903
; Sequence 6903, Application US/10027632
```

```

; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6903
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-6903

Query Match      82.0%; Score 16.4; DB 5; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  3 AGTAAATGCAGAAACAGG 20
Db   ||||| ||||| ||||| |||||
    249 AGTAAATGGAGAAACAGG 266

RESULT 53
US-10-027-632-324576
; Sequence 324576, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324576
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324576

Query Match      82.0%; Score 16.4; DB 5; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  3 AGTAAATGCAGAAACAGG 20
Db   ||||| ||||| ||||| |||||
    249 AGTAAATGGAGAAACAGG 266

RESULT 54
US-10-027-632-6903
; Sequence 6903, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6903
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-6903

Query Match      82.0%; Score 16.4; DB 6; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  3 AGTAAATGCAGAAACAGG 20
Db   ||||| ||||| ||||| |||||
    249 AGTAAATGGAGAAACAGG 266

RESULT 55
US-10-027-632-324576
; Sequence 324576, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324576
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324576

Query Match      82.0%; Score 16.4; DB 5; Length 504;
```



; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 324576  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-324576

Query Match 82.0%; Score 16.4; DB 6; Length 504;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAGG 20  
|||||  
Db 249 AGTAAATGCAGAAACAGG 266

RESULT 56  
US-10-472-928-4807/C  
; Sequence 4807, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026926WO  
; CURRENT APPLICATION NUMBER: US/10/472,928  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: GB-0107658.7  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: Seqwin99, version 1.03  
; SEQ ID NO 4807  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-472-928-4807

Query Match 82.0%; Score 16.4; DB 8; Length 507;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 501 AAGTAAATGCAGTAAACAG 484

RESULT 57  
US-10-029-386-8721  
; Sequence 8721, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: ABOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 8721  
; LENGTH: 512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011311.11  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
; OTHER INFORMATION: SWISSPROT HIT: P80665, EVALUE 4.80e+00  
; OTHER INFORMATION: EST HUMAN HIT: A1201312.1, EVALUE 8.00e-07  
; OTHER INFORMATION: NT HIT: gi14729768, EVALUE 2.00e-06  
US-10-029-386-8721

Query Match 82.0%; Score 16.4; DB 6; Length 512;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 331 AAGCAATGCAGAAACAG 348

RESULT 58  
US-10-027-632-50588  
; Sequence 50588, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 50588  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-50588

Query Match 82.0%; Score 16.4; DB 5; Length 545;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 438 AAGTAAATGCAGAAATCAG 455

RESULT 59  
US-10-027-632-50589  
; Sequence 50589, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20

Search completed: January 28, 2006, 01:45:44  
Job time : 350.966 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 201.573 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 60 summaries

Database :

- Published Applications NA New.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	2866	7 US-10-750-185-39068	Sequence 39068, A
2	16.8	84.0	2866	7 US-10-750-623-39068	Sequence 39068, A
3	16.4	82.0	4320	7 US-10-750-185-27669	Sequence 27669, A
4	16.4	82.0	4320	7 US-10-750-623-27669	Sequence 27669, A
5	16.4	82.0	73072	8 US-11-124-368A-2919	Sequence 2919, Ap
6	16.4	82.0	149382	7 US-10-995-561-13272	Sequence 13272, A
7	15.8	79.0	201	7 US-10-995-561-43569	Sequence 43569, A
8	15.8	79.0	604	7 US-10-980-388-44	Sequence 44, Appl
9	15.8	79.0	954	7 US-10-750-185-46466	Sequence 46466, A
10	15.8	79.0	954	7 US-10-750-623-46466	Sequence 46466, A
11	15.8	79.0	1108	7 US-10-750-185-45741	Sequence 45741, A
12	15.8	79.0	1108	7 US-10-750-623-45741	Sequence 45741, A
13	15.8	79.0	1223	7 US-10-750-185-56170	Sequence 56170, A
14	15.8	79.0	1223	7 US-10-750-623-56170	Sequence 56170, A
15	15.8	79.0	1435	7 US-10-750-185-52569	Sequence 52569, A
16	15.8	79.0	1435	7 US-10-750-623-52569	Sequence 52569, A
17	15.8	79.0	1447	7 US-10-750-185-61366	Sequence 61366, A
18	15.8	79.0	1447	7 US-10-750-623-61366	Sequence 61366, A
19	15.8	79.0	1770	8 US-11-128-061-572	Sequence 572, Appl
20	15.8	79.0	1770	8 US-11-128-049-572	Sequence 572, Appl
21	15.8	79.0	2659	7 US-10-750-185-29929	Sequence 29929, A
22	15.8	79.0	2659	7 US-10-750-623-29929	Sequence 29929, A

c	23	15.8	79.0	3001	8 US-11-145-703-194	Sequence 194, App
c	24	15.8	79.0	21623	7 US-10-995-561-13497	Sequence 13497, A
c	25	15.8	79.0	44362	7 US-10-995-561-13422	Sequence 13422, A
c	26	15.8	79.0	119160	8 US-11-121-086-12	Sequence 12, Appl
c	27	15.8	79.0	156544	8 US-11-121-086-81	Sequence 81, Appl
c	28	15.8	79.0	159695	8 US-11-121-086-56	Sequence 56, Appl
c	29	15.8	79.0	394468	7 US-10-995-561-13473	Sequence 13473, A
c	30	15.8	79.0	1125000	7 US-10-995-561-13286	Sequence 13286, A
c	31	15.4	77.0	598	7 US-10-750-185-19994	Sequence 19994, A
c	32	15.4	77.0	598	7 US-10-750-623-19994	Sequence 19994, A
c	33	15.4	77.0	881	7 US-10-750-185-47634	Sequence 47634, A
c	34	15.4	77.0	881	7 US-10-750-623-47634	Sequence 47634, A
c	35	15.4	77.0	884	7 US-10-750-185-62704	Sequence 62704, A
c	36	15.4	77.0	884	7 US-10-750-623-62704	Sequence 62704, A
c	37	15.4	77.0	1401	7 US-10-750-185-56765	Sequence 56765, A
c	38	15.4	77.0	1401	7 US-10-750-623-56765	Sequence 56765, A
c	39	15.4	77.0	1423	7 US-10-750-185-41820	Sequence 41820, A
c	40	15.4	77.0	1423	7 US-10-750-623-41820	Sequence 41820, A
c	41	15.4	77.0	2343	7 US-10-750-185-42759	Sequence 42759, A
c	42	15.4	77.0	2343	7 US-10-750-623-42759	Sequence 42759, A
c	43	15.4	77.0	2500	7 US-10-750-185-36275	Sequence 36275, A
c	44	15.4	77.0	2500	7 US-10-750-623-36275	Sequence 36275, A
c	45	15.4	77.0	185393	8 US-11-121-086-101	Sequence 101, Appl
c	46	15.4	77.0	319608	8 US-11-145-703-1	Sequence 1, Appl
c	47	15.4	77.0	319608	8 US-11-145-703-1	Sequence 1, Appl
c	48	15.4	77.0	1082144	8 US-11-117-187-211	Sequence 211, App
c	49	15.2	76.0	25	8 US-11-136-527-326267	Sequence 326267, A
c	50	15.2	76.0	201	7 US-10-995-561-63705	Sequence 63705, A
c	51	15.2	76.0	201	7 US-11-124-368A-16091	Sequence 16091, A
c	52	15.2	76.0	201	8 US-11-124-368A-16143	Sequence 16143, A
c	53	15.2	76.0	579	8 US-11-128-061-1990	Sequence 1990, Ap
c	54	15.2	76.0	579	8 US-11-128-061-5632	Sequence 5632, Ap
c	55	15.2	76.0	579	8 US-11-128-049-1990	Sequence 1990, Ap
c	56	15.2	76.0	579	8 US-11-128-049-5632	Sequence 5632, Ap
c	57	15.2	76.0	600	8 US-11-136-527-6925	Sequence 6925, Ap
c	58	15.2	76.0	1115	7 US-10-750-185-28011	Sequence 28011, A
c	59	15.2	76.0	1115	7 US-10-750-623-28011	Sequence 28011, A
c	60	15.2	76.0	1538	7 US-10-750-185-42218	Sequence 42218, A

ALIGNMENTS

RESULT 1  
US-10-750-185-39068  
; Sequence 39068, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39068  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Bovine 1986680837502  
US-10-750-185-39068

Query Match 84.0%; Score 16.8; DB 7; Length 2866;  
Best Local Similarity 90.0%; Pred. No. 1.se+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||  
Db 235 CAAGTAAATGCAGCCACAGG 254

RESULT 2  
US-10-750-623-39068  
; Sequence 39068, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39068  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-623-39068

Query Match 84.0%; Score 16.8; DB 7; Length 2866;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||  
Db 235 CAAGTAAATGCAGCCACAGG 254

RESULT 3  
US-10-750-185-27669/c  
; Sequence 27669, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27669  
; LENGTH: 4320  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-185-27669

Query Match 82.0%; Score 16.4; DB 7; Length 4320;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 18  
|||||  
Db 4276 CAAGTAAATGCAGTAACA 4259

RESULT 4  
US-10-750-623-27669/c  
; Sequence 27669, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27669  
; LENGTH: 4320  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-623-27669

Query Match 82.0%; Score 16.4; DB 7; Length 4320;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 18  
|||||  
Db 4276 CAAGTAAATGCAGTAACA 4259

RESULT 5  
US-11-124-368A-2919  
; Sequence 2919, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2919  
; LENGTH: 73072  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2919

Query Match 82.0%; Score 16.4; DB 8; Length 73072;  
Best Local Similarity 94.4%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAACAGG 20  
|||||  
Db 38284 AGAATGCAGAACAGG 38301

RESULT 6  
US-10-995-561-13272

```
; Sequence 13272, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13272
; LENGTH: 149382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(149382)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13272

Query Match      82.0%; Score 16.4; DB 7; Length 149382;
Best Local Similarity 94.4%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      |||||
Db      94792 AAGTAAATGCAGAAACAG 94809

RESULT 7
US-10-995-561-43569/c
; Sequence 43569, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43569
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-43569

Query Match      79.0%; Score 15.8; DB 7; Length 201;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
      |||||
Db      197 CGAATTAATGCAGAAACAG 179

RESULT 8
US-10-980-388-44/c
; Sequence 44, Application US/10980388
; Publication No. US20050255490A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Parodi, Luis A.
; APPLICANT: Hiebsch, Ronald R.
; APPLICANT: Lind, Peter
; APPLICANT: Kaytes, Paul S.
; APPLICANT: Ruff, Valerie
; APPLICANT: Huff, Rita M.
; APPLICANT: Wood, Linda S.
```

```
; TITLE OF INVENTION: Novel G Protein-Coupled Receptors Cross-Reference To Related Appl
; FILE REFERENCE: 00325.US1
; CURRENT APPLICATION NUMBER: US/10/980,388
; CURRENT FILING DATE: 2004-11-02
; PRIOR APPLICATION NUMBER: US/09/791,932
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,305
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,304
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,303
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,397
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,247
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/188,880
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/217,369
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/217,370
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/218,492
; PRIOR FILING DATE: 2000-07-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-980-388-44

Query Match      79.0%; Score 15.8; DB 7; Length 604;
Best Local Similarity 89.5%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
      |||||
Db      94 CAAGTAAATGCAGAAACAG 76

RESULT 9
US-10-750-185-46466
; Sequence 46466, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MW1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46466
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Bovine 19866880562987
US-10-750-185-46466

Query Match      79.0%; Score 15.8; DB 7; Length 954;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 20
```

```
Db      551 AAGTAAATTCAGAACAG 569
|||||
RESULT 10
US-10-750-623-46466
; Sequence 46466, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 46466
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Bovine 19866880562987
US-10-750-623-46466

Query Match      79.0%; Score 15.8; DB 7; Length 954;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAACAG 20
|||||
Db      551 AAGTAAATTCAGAACAG 569
|||||

RESULT 11
US-10-750-185-45741/c
; Sequence 45741, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45741
; LENGTH: 1108
; TYPE: DNA
; ORGANISM: Bovine 19866880792232
US-10-750-185-45741

Query Match      79.0%; Score 15.8; DB 7; Length 1108;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAACAG 19
|||||
Db      250 CAAGAAATGCAGACACAG 232
|||||

RESULT 12
US-10-750-623-45741/c
; Sequence 45741, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45741
; LENGTH: 1108
; TYPE: DNA
; ORGANISM: Bovine 19866880792232
US-10-750-623-45741

Query Match      79.0%; Score 15.8; DB 7; Length 1108;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAACAG 19
|||||
Db      250 CAAGAAATGCAGACACAG 232
|||||

RESULT 13
US-10-750-185-56170/c
; Sequence 56170, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56170
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Bovine 19866880790313
US-10-750-185-56170

Query Match      79.0%; Score 15.8; DB 7; Length 1223;
Best Local Similarity 89.5%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAACAG 19
|||||
Db      289 CAATAAATGCAGATACAG 271
|||||

RESULT 14
US-10-750-623-56170/c
```



```
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2003-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61366
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Bovine 19866880352244
US-10-750-623-61366

Query Match          79.0%; Score 15.8; DB 7; Length 1447;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 1008 CAGGAAATGCAGAAACAG 990

RESULT 19
US-11-128-061-472
; Sequence 572, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 20
US-11-128-049-572
; Sequence 572, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 21
US-10-750-185-29929
; Sequence 29929, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29929
; LENGTH: 2659
; TYPE: DNA
; ORGANISM: Bovine 19866880628714
US-10-750-185-29929

Query Match          79.0%; Score 15.8; DB 7; Length 2659;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 601 CAAGTAGATGGAGAAACAG 619

RESULT 22
US-10-750-623-29929
```

```
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 21
US-10-750-185-29929
; Sequence 29929, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29929
; LENGTH: 2659
; TYPE: DNA
; ORGANISM: Bovine 19866880628714
US-10-750-185-29929

Query Match          79.0%; Score 15.8; DB 7; Length 2659;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 601 CAAGTAGATGGAGAAACAG 619

RESULT 22
US-10-750-623-29929
```



```
; Sequence 29929, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 29929
; LENGTH: 2659
; TYPE: DNA
; ORGANISM: Bovine 1986680628714
US-10-750-623-29929

Query Match          79.0%; Score 15.8; DB 7; Length 2659;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 601 CAAGTAGATGCAGAAACAG 619

RESULT 23
US-11-145-703-194/c
; Sequence 194, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US16.DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.pm
; SEQ ID NO 194
; LENGTH: 3001
```

```
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1501
; OTHER INFORMATION: 99-26223-225 : polymorphic base G or T
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1481..1500
; OTHER INFORMATION: 99-26223-225.mis1,
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1502..1520
; OTHER INFORMATION: 99-26223-225.mis2, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1277..1297
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1842..1862
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-26223-225 probe
US-11-145-703-194

Query Match          79.0%; Score 15.8; DB 8; Length 3001;
Best Local Similarity 89.5%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 1917 CAATTATATGCAGAAACAG 1899

RESULT 24
US-10-995-561-13497
; Sequence 13497, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13497
; LENGTH: 21623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13497

Query Match          79.0%; Score 15.8; DB 7; Length 21623;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 17255 CAATTATATGCAGAAACAG 17273

RESULT 25
US-10-995-561-13422/c
; Sequence 13422, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13422
; LENGTH: 44362
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(44362)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13422

Query Match          79.0%; Score 15.8; DB 7; Length 44362;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 11504 CAATTAATGCAGAAACAG 11486

RESULT 26
US-11-121-086-12/c
; Sequence 12, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 12
; LENGTH: 119160
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-12

Query Match          79.0%; Score 15.8; DB 8; Length 119160;
Best Local Similarity 89.5%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 80153 CAATTAATGCAGAAACAG 80135

RESULT 27
US-11-121-086-81/c
; Sequence 81, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 81
; LENGTH: 156544
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-81

Query Match          79.0%; Score 15.8; DB 8; Length 156544;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 68402 CAATCAATGCAGAAACAG 68384

RESULT 28
US-11-121-086-56/c
; Sequence 56, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 56
; LENGTH: 159695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-56

Query Match          79.0%; Score 15.8; DB 8; Length 159695;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 80674 CAATTAATGCAGAAACAG 80656

RESULT 29
US-10-995-561-13473/c
; Sequence 13473, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13473
; LENGTH: 394468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(394468)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13473

Query Match          79.0%; Score 15.8; DB 7; Length 394468;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 20
   ||||| ||||| ||||| |||||
```

Db 171081 AAGTAAATGCTAAACAGG 171063

```
RESULT 30
US-10-995-561-13286/c
; Sequence 13286, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13286
; LENGTH: 1125000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1125000)
; OTHER INFORMATION: n is any nucleotide
US-10-995-561-13286
```

Query Match 79.0%; Score 15.8; DB 7; Length 1125000;  
Best Local Similarity 89.5%; Pred. No. 5e+02; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2;

QY 1 CAAGTAAATGCAGAAACAG 19  
Db 657182 CGAATAATGCAGAAACAG 657164

```
RESULT 31
US-10-750-185-19994
; Sequence 19994, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19994
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Bovine MMBT01307
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(101)
; OTHER INFORMATION: n is any nucleotide
US-10-750-185-19994
```

Query Match 77.0%; Score 15.4; DB 7; Length 598;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02; Indels 0; Gaps 0;  
Matches 16; Conservative 0; Mismatches 1;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323

```
RESULT 32
US-10-750-623-19994
; Sequence 19994, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19994
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Bovine MMBT01307
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(101)
; OTHER INFORMATION: n is any nucleotide
US-10-750-623-19994
```

Query Match 77.0%; Score 15.4; DB 7; Length 598;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02; Indels 0; Gaps 0;  
Matches 16; Conservative 0; Mismatches 1;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323

```
RESULT 33
US-10-750-185-47634/c
; Sequence 47634, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47634
; LENGTH: 881
; TYPE: DNA
; ORGANISM: Bovine 19866881682228
US-10-750-185-47634
```

Query Match 77.0%; Score 15.4; DB 7; Length 881;  
Best Local Similarity 94.1%; Pred. No. 5.4e+02; Indels 0; Gaps 0;  
Matches 16; Conservative 0; Mismatches 1;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323



```
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56765
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Bovine 19866881596118
US-10-750-623-56765

Query Match      77.0%; Score 15.4; DB 7; Length 1401;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACA 18
Db      474 AAGTAAATGCAGAAACA 490

RESULT 39
US-10-750-185-41820
; Sequence 41820, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41820
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Bovine 19866881002825
US-10-750-185-41820

Query Match      77.0%; Score 15.4; DB 7; Length 1423;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACA 18
Db      1108 AAGTAAATGCAGAAACA 1124

RESULT 40
US-10-750-623-41820
; Sequence 41820, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```

```
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41820
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Bovine 19866881002825
US-10-750-623-41820

Query Match      77.0%; Score 15.4; DB 7; Length 1423;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACA 18
Db      1108 AAGTAAATGCAGAAACA 1124

RESULT 41
US-10-750-185-42759
; Sequence 42759, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42759
; LENGTH: 2343
; TYPE: DNA
; ORGANISM: Bovine 19866880345005
US-10-750-185-42759

Query Match      77.0%; Score 15.4; DB 7; Length 2343;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAAC 17
Db      1301 CAAGTAAATGCAGAAAC 1317

RESULT 42
US-10-750-623-42759
; Sequence 42759, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```

```
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; CURRENT FILING DATE: 2003-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42759
; TYPE: DNA
; LENGTH: 2343
; ORGANISM: Bovine 19866880345005
US-10-750-623-42759

Query Match          77.0%; Score 15.4; DB 7; Length 2343;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAAC 17
   ||||| ||||| |||||
Db 1301 CAAGTAAATGCAGAAAC 1317

RESULT 43
US-10-750-185-46275
; Sequence 36275, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36275
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: Bovine 19866880753661
US-10-750-185-36275

Query Match          77.0%; Score 15.4; DB 7; Length 2500;
Best Local Similarity 94.1%; Pred. No. 6.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 206 AGTAAATGCAGAAACAG 222

RESULT 44
US-10-750-623-36275
; Sequence 36275, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

```
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36275
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: Bovine 19866880753661
US-10-750-623-36275

Query Match          77.0%; Score 15.4; DB 7; Length 2500;
Best Local Similarity 94.1%; Pred. No. 6.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 206 AGTAAATGCAGAAACAG 222

RESULT 45
US-11-121-086-101
; Sequence 101, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIN version 3.3
; SEQ ID NO 101
; LENGTH: 185393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-101

Query Match          77.0%; Score 15.4; DB 8; Length 185393;
Best Local Similarity 94.1%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 12409 AGTAAATGCAGAAACAG 12425

RESULT 46
US-11-145-703-1
; Sequence 1, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihaïn, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53 US16 DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
```

```

; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239807
; OTHER INFORMATION: exon N2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239853
; OTHER INFORMATION: exon N complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240569
; OTHER INFORMATION: exon M1117 complement g34872 gene
```

FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240528..240596 ;  
OTHER INFORMATION: exon M1090 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240528..240617 ;  
OTHER INFORMATION: exon M1069 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240528..240644 ;  
OTHER INFORMATION: exon M52 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240528..240824 ;  
OTHER INFORMATION: exon M862 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240528..240994 ;  
OTHER INFORMATION: exon M692 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: exon ;  
LOCATION: 240800..240993 ;  
OTHER INFORMATION: exon M51 complement g34872 gene ;  
FEATURE: ;  
NAME/KEY: misc feature ;  
LOCATION: 241686..243685 ;  
OTHER INFORMATION: 5'regulatory region g34872 gene ;  
FEATURE: ;  
NAME/KEY: misc feature ;  
LOCATION: 290652..292652 ;  
OTHER INFORMATION: 3'regulatory region g34665 gene ;

Query Match 77.0%; Score 15.4; DB 8; Length 319608;  
Best Local Similarity 94.1%; Pred. No. 8.e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 AAGTAAATGCAGAAACA 18  
|||||  
Db 161192 AAGTGAATGCAGAAACA 161208

RESULT 47  
US-11-145-703-1/c  
Sequence 1, Application US/11145703  
Publication No. US20050260667A1  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bouqueleret, Lydie  
APPLICANT: Bihain, Bernard  
APPLICANT: Essioux, Laurent  
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS  
FILE REFERENCE: 53.US16.DIV  
CURRENT APPLICATION NUMBER: US/11/145,703  
CURRENT FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US/10/147,603  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: 09/539,333  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/126,903  
PRIOR FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: US 60/131,971  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/132,065  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/143,928  
PRIOR FILING DATE: 1999-07-14

PRIOR APPLICATION NUMBER: US 60/145,915  
PRIOR FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: US 60/146,453  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/146,452  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/162,288  
PRIOR FILING DATE: 1999-10-28  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 234  
SOFTWARE: Patent.Pm  
SEQ ID NO 1  
LENGTH: 319608  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 31..1107  
OTHER INFORMATION: 5'regulatory region g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1108..1289  
OTHER INFORMATION: exon A g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 14877..14920  
OTHER INFORMATION: exon B g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 18778..18862  
OTHER INFORMATION: exon Bbis g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 25593..25740  
OTHER INFORMATION: exon C g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 29388..29502  
OTHER INFORMATION: exon D g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 29967..30282  
OTHER INFORMATION: exon E g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 64666..64812  
OTHER INFORMATION: exon F g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 65505..65853  
OTHER INFORMATION: exon G g35018 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 65854..67854  
OTHER INFORMATION: 3'regulatory region g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 94124..94964  
OTHER INFORMATION: exon g35017  
FEATURE:  
NAME/KEY: exon  
LOCATION: 201188..201234  
OTHER INFORMATION: exon S g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 214676..214793  
OTHER INFORMATION: exon T g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215702..215746  
OTHER INFORMATION: exon U g35030 gene  
FEATURE:  
NAME/KEY: exon



LOCATION: 216836..216915  
OTHER INFORMATION: exon V g35030 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 213818..215818  
OTHER INFORMATION: 3'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215941  
OTHER INFORMATION: exon R complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215975  
OTHER INFORMATION: exon Rbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..216952  
OTHER INFORMATION: exon Qbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..217061  
OTHER INFORMATION: exon Q complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 217027..217061  
OTHER INFORMATION: exon Q1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 229647..229742  
OTHER INFORMATION: exon X complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 230408..230721  
OTHER INFORMATION: exon P complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231272..231412  
OTHER INFORMATION: exon Obis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231787..231880  
OTHER INFORMATION: exon O2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231870..231879  
OTHER INFORMATION: exon O1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 234174..234321  
OTHER INFORMATION: exon O complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237406..237428  
OTHER INFORMATION: exon Nbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239807  
OTHER INFORMATION: exon N2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239853  
OTHER INFORMATION: exon N complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240569  
OTHER INFORMATION: exon M1117 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240596  
OTHER INFORMATION: exon M1090 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240617

OTHER INFORMATION: exon M1069 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240644  
OTHER INFORMATION: exon MS2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240824  
OTHER INFORMATION: exon M862 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240994  
OTHER INFORMATION: exon M692 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..241685  
OTHER INFORMATION: exon M1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240800..240993  
OTHER INFORMATION: exon MS1 complement g34872 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 241686..243685  
OTHER INFORMATION: 5'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 290652..292652  
OTHER INFORMATION: 3'regulatory region g34665 gene

Query Match 77.0%; Score 15.4; DB 8; Length 319608;  
Best Local Similarity 94.1%; Pred. No. 8.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTAAATGCAGAAACAGG 20  
|||||

DB 193336 GTAAATGTAGAAACAGG 193320  
|||||

RESULT 48

US-11-117-187-211  
; Sequence 211, Application US/11117187  
; Publication No. US20050266560A1

GENERAL INFORMATION:

APPLICANT: PREUSS, DAPHNE  
APPLICANT: COPENHAVER, GREGORY  
TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS

FILE REFERENCE: ARCD:309US

CURRENT APPLICATION NUMBER: US/11/117,187

CURRENT FILING DATE: 2005-04-28

PRIOR APPLICATION NUMBER: US/09/531,120

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/125,219

PRIOR FILING DATE: 1999-03-18

NUMBER OF SEQ ID NOS: 212

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 211

LENGTH: 1082144

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-11-117-187-211

Query Match 77.0%; Score 15.4; DB 8; Length 1082144;

Best Local Similarity 94.1%; Pred. No. 6.1e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAAC 17  
|||||

DB 190955 CAAGTAAATACAGAAAC 190971  
|||||

RESULT 49

US-11-136-527-326267

```
; Sequence 326267, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 326267
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-326267

Query Match 76.0%; Score 15.2; DB 8; Length 25;
Best Local Similarity 85.0%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 1 CAAGTAAATGCAGAAACCG 20

RESULT 50
US-10-995-561-63705/c
; Sequence 63705, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63705
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-63705

Query Match 76.0%; Score 15.2; DB 7; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 161 CAAGTAAATGCAGAAACAGG 142

RESULT 51
US-11-124-368A-16091
; Sequence 16091, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
```

```
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16091
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-16091

Query Match 76.0%; Score 15.2; DB 8; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 71 CAAGCAATGCTGAAACAGG 90

RESULT 52
US-11-124-368A-16143
; Sequence 16143, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16143
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-16143

Query Match 76.0%; Score 15.2; DB 8; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 121 CAAGCAATGCTGAAACAGG 140

RESULT 53
US-11-128-061-1990
; Sequence 1990, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
```

; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1990  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-061-1990

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
Db 539 CAAGTCAATGCAGAAATCATG 558

RESULT 54

US-11-128-061-5632  
; Sequence 5632, Application US/11128061  
; Publication No. US20060003958A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS  
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION  
; FILE REFERENCE: 01997.027701  
; CURRENT APPLICATION NUMBER: US/11/128,061  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5632  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-061-5632

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
Db 539 CAAGTCAATGCAGAAATCATG 558

RESULT 55

US-11-128-049-1990  
; Sequence 1990, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11

; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1990  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-049-1990

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
Db 539 CAAGTCAATGCAGAAATCATG 558

RESULT 56

US-11-128-049-5632  
; Sequence 5632, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FO  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5632  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-049-5632

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
Db 539 CAAGTCAATGCAGAAATCATG 558

RESULT 57

US-11-136-527-6925  
; Sequence 6925, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M.  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6925  
; LENGTH: 600

```

; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6925

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 8; Length 600;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
DB 527 CAAGTAAATGCAGAAACCGG 546

RESULT 58
US-10-750-185-28011
; Sequence 28011, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28011
; LENGTH: 1115
; TYPE: DNA
; ORGANISM: Bovine 19866880751726
US-10-750-185-28011

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1115;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
DB 1083 CAAATAAATGGGAAACAGG 1102

RESULT 59
US-10-750-623-28011
; Sequence 28011, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28011
; LENGTH: 1115
; TYPE: DNA
; ORGANISM: Bovine 19866880751726
US-10-750-623-28011

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1115;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
DB 1083 CAAATAAATGGGAAACAGG 1102

RESULT 60
US-10-750-185-42218/c
; Sequence 42218, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42218
; LENGTH: 1538
; TYPE: DNA
; ORGANISM: Bovine 19866880881480
US-10-750-185-42218

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1538;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
DB 443 CAAGTCAAAACAGAAACAGG 424

Search completed: January 28, 2006, 02:00:54
Job time : 208.823 secs
```